

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 6, 2005, 05:58:22 ; Search time 327 Seconds
(without alignments)
9157.145 Million cell updates/sec

Title: US-10-068-870A-4
Perfect score: 1830
Sequence: 1 atgaaaaaagaatttagc.....tgaagatttagatg 1830

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTCUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95.6	5.2	5361	3	US-08-973-462-2
2	95.6	5.2	6152	3	US-08-973-462-1
3	87	4.8	3095	6	5231168-1
4	87	4.8	3095	6	5231168-1
5	85.2	4.7	8155	4	US-08-956-171E-63
6	85.2	4.7	8155	4	US-08-781-986A-63
7	82.2	4.5	2370	4	US-09-754-947-2
8	76	4.2	1039	4	US-09-902-540-1280
9	73.4	4.0	4766	5	PCT-US93-07261-10
10	71.6	3.9	14066	4	US-09-601-198-56
11	70	3.8	3279	3	US-08-446-137B-1
12	69.8	3.8	612	4	US-09-902-540-1357
13	69.4	3.8	1891	3	US-08-973-462-3
14	69.4	3.8	2489	3	US-09-141-047-7
15	69.4	3.8	9636	1	US-08-323-170B-1
16	69.4	3.8	9636	3	US-08-954-441-1
17	68.4	3.7	7218	1	US-08-232-463-14
18	68.4	3.7	187169	4	US-09-949-016-12776
19	68.4	3.7	191569	4	US-09-949-016-15940
20	67.8	3.7	3057	4	US-09-601-198-55
21	65	3.6	3356	3	US-09-379-523-4
22	65	3.6	580073	4	US-08-545-528D-1
23	64.4	3.5	1431	3	US-09-316-083-2
24	64.4	3.5	1431	4	US-09-933-700-2
25	63.6	3.5	1956	3	US-08-559-896B-1
26	63.6	3.5	1956	4	US-09-351-794A-1
27	63	3.4	19124	2	US-08-487-826B-13

c	28	62.8	3.4	3543	4	US-09-710-279-4149	Sequence 4149, Ap
	29	62.8	3.4	3840	4	US-09-710-279-3187	Sequence 3187, Ap
	30	62.8	3.4	11091	3	US-09-134-001C-2243	Sequence 2243, Ap
	31	62.8	3.4	205044	4	US-09-949-016-15851	Sequence 15851, A
	32	62.8	3.4	205044	4	US-09-949-016-15852	Sequence 15852, A
	33	62.8	3.4	205044	4	US-09-949-016-15853	Sequence 15853, A
	34	62.8	3.4	223471	4	US-09-949-016-12387	Sequence 12387, A
	35	62.8	3.4	223471	4	US-09-949-016-12724	Sequence 12724, A
	36	62.8	3.4	223471	4	US-09-949-016-12725	Sequence 12725, A
	37	62.6	3.4	3489	2	US-08-728-323A-1	Sequence 1, Appli
	38	62.6	3.4	3489	3	US-09-298-568-1	Sequence 1, Appli
	39	62.6	3.4	3489	4	US-09-410-399-1	Sequence 1, Appli
	40	62.6	3.4	3489	4	US-09-894-273-1	Sequence 1, Appli
c	41	62.6	3.4	32207	2	US-08-770-379-20	Sequence 20, Appl
	42	62.6	3.4	32207	3	US-08-757-669A-20	Sequence 20, Appl
c	43	62.6	3.4	32207	3	US-09-230-371A-20	Sequence 31, Appl
	44	61.4	3.4	810	4	US-09-601-198-31	Sequence 30531, A
c	45	60.8	3.3	601	4	US-09-949-016-30531	

ALIGNMENTS

RESULT 1

US-08-973-462-2
; Sequence 2, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5361
; TYPE: DNA
; ORGANISM: P. falciparum
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (1)..(5361)
; US-08-973-462-2

Query Match	5.2%	Score 95.6;	DB 3;	Length 5361;
Best Local Similarity	43.6%	Pred. No. 6e-11;		
Matches	645;	Conservative	0;	Mismatches 809;
				Indels 24;
				Gaps 4;
Qy	233	CAATGAGAAGATCTTCAACTCGGAGATATAAAGTTTATTGAGAACAACTT	292	
Db	770	CAAGTGTGAAGAAGATATAGCTTCAAGTGTGAAAGTATAGATTCAAGTATTGAAG	829	
Qy	293	TAACACAGCTAATGGAATGAAGATATATGAAGCAA---CTTTAAAAATTTAGATG	349	
Db	830	AAATGTAGCTCAACTGTGTAAGAAATCGTAGCTCCAGTGTGTAAGAAGTGTGCTC	889	
Qy	350	CAGGAGAATATGCTATTATATAGATTAACTTTATATATGCTAAACTGTTGAAATTAAG	409	
Db	890	CAAGTGTGAAGAAGTGTAGAGAAATCTTGAAGAAGTGTAGCTGAAAATGTTGAAG	949	
Qy	410	TAGTAGCAGCTAGTGAAGAAACAGTAGTGTATCTAGTATGCGAAAATAGTCAAAAG	469	
Db	950	AAAGTGTAGCTGAAAATGTTGAAGAAGTGTAGCTGAAAATGTTGAAGAAGTGTAGCTG	1009	
Qy	470	ATATAGCTGAAAATATGTTGTTGAGACAAAGCTTTAGAAATGCACTAAACTATAA	529	
Db	1010	AAATGTGAAGAATCGTAGCTCAACTGTTGAAGAATCGTAGCTCCAACTGTTGAAG	1069	

QY 530 ATGCTCAGATTTTCAGTAAACTGATAGTTACTATCAAGTAGTCTCTTTATCCAAAGGAA 589
 Db 1070 AAATGTAGCTCCAAGTGTGTGAAGAGTGTGGCTCCAAGTGTGTGAAGAAAGTGTAGAAG 1129
 QY 590 AGAGATTACAAAGTTTCTCAACTTATAGAGCTACAAATATATAATGAAGCAACTGCAATAG 649
 Db 1130 AAAATGTTGAAGAAAGTAGCTGAANAATGTTGAAGAAAGTAGCTGAANAATGTTGAAG 1189
 QY 650 GTAATACACAGTAATTAATCTCTAAATCTACTAGTAAGAGTAATTTAAAGACTGCAG 709
 Db 1190 AAAGTGTAGCTGAAAATGTTGAAGAAAGTAGCTGAAAATGTTGAAGAAAGTAGCTG 1249
 QY 710 TAGAAGAGTTACAAAATTTGAATGCTAGTTATCTTAATACTACAACTTTAGCTGGTGAG 769
 Db 1250 AAAATGTTGAAGAAATCGTAGCTCCAACCTGTTGAAGAAATCGTAGCTCCAACCTGTTGAAG 1309
 QY 770 ACAGAAATACAAAAGCTATAGAGTAAGTAAGAAATATTAATAATATGATGCGCAAAAT 829
 Db 1310 AAATGTAGCTCCAAGTGTGTGAAGAAAGTGTGGCTCCAAGTGTGTGAAGAAAGTAGAG 1369
 QY 830 CAGATCAATCAGCTGATGTTAAAGAGAAATGTTAAATAATGTTGATTAAGTAGTGCAATG 889
 Db 1370 AAAATGTTGAAGAAAGTAGCTGAAAATGTTGAAGAAAGTAGCTGAAAATGTTGAAG 1429
 QY 890 CACTAGTAGATGATTTAGTTGGCTCCTTTAGCAGCAAGAAAGATGCTCCACTATTAT 949
 Db 1430 AAAGTGTAGCTGAAAATGTTGAAGAAAGTAGCTGAAAATGTTGAAGAAAGTAGTAGCTG 1489
 QY 950 TAACTTCAAAAGATAAATTAGATTCGTCAGTAAA-----ATCTGAAATAAAGAGAG 1000
 Db 1490 AAAATGTTGAAGAAAGTAGCTGAAAATGTTGAAGAAAGTAGCTGAAAATGTTGAAG 1549
 QY 1001 TTTTAGACTTAAAACTTCAACAGAAAGTACAGGAAAAACAGTTTATATAGCTGGTGGAG 1060
 Db 1550 AAAATGTTGAAGAAAGTAGCTGAAAATGTTGAAGAAAGTAGCTGAAAATGTTGAAG 1609
 QY 1061 TTAATAGTCTATCTAAGAAAGTGTGAAGAAATGAGAAATCAATGGATTTAAAGTTGAA 1120
 Db 1610 CAAAGTGTGTGAAGAAAGTGGCTCCAAGTGTGTGAAGAAAGTAGAGAAATGTTGAAG 1669
 QY 1121 GATTTCTCAGTGATGATGATGAAATCTTTTAAATAATAGCAGGTGAAATAGGCTTAG 1180
 Db 1670 AAAGTGTAGCTGAAAATGTTGAAGAAAGTAGCTGAAAATGTTGAAGAAAGTAGCTG 1729
 QY 1181 ATAATGATTAAGCTTATAGTTGGTGGAAACAGGATTAAGAGTCCATGAGTATAGCTT 1240
 Db 1730 AAAATGTTGAAGAAAGTAGCTGAAAATGTTGAAGAAATCGTAGCTCCAACCTGTTGAAG 1789
 QY 1241 CAGTTGCTTCTACTAAATTTAGATGGTAATG-----GTGTTGTAGATAGAACAAATG 1291
 Db 1790 AAATCGTAGCTCCAATGTTGAAGAAATGTAGCTCCAAGTGTGTGAAGAAAGTAGGCTC 1849
 QY 1292 GACATGCTACTCCAATAGTTGTTAGATGGAAGAAAGCTGATAAATAATCTGATGACTTAG 1351
 Db 1850 CAAAGTGTGAAGAAAGTAGAGAAATGTTGAAGAAAGTAGCTGAAAATGTTGAAG 1909
 QY 1352 ATAGTTTCTTAGAAGCGCTGATGATGATATAATAGTGGATTTGCAAGTGTATCTGAAA 1411
 Db 1910 AAAGTGTAGCTGAAAATGTTGAAGAAAGTAGCTGAAAATGTTGAAGAAATCGTAGCTC 1969
 QY 1412 AGATGGAAGAGCTATATCAGATGCTGCTGTAAGGCGGTACAGAGTTAAAGCGGAG 1471
 Db 1970 CAACTGTTGAAGAAATCGTAGCTCCAACCTGTTGAAGAAATGTAGCTCCAAGTGTGTAG 2029
 QY 1472 ATAGACACAGACACTACTCTGAAGTTTATAAAACATATTTATGCTAATGATCTGAAAATAG 1531
 Db 2030 AAAGTGTGCTCCAAGTGTGAAGAAAGTAGAGAAATGTTGAAGAAAGTGTGAAGAAAGT 2086
 QY 1532 CTAAGCTGAGCTTTTAGATAAAGATTCAGGTGCTTCAAGTAGTGAATGACGAGAGTATTTA 1591
 Db 2087 CTGAAAATGTTGAAGAAAGTAGCTGAAAATGTTGAAGAAAGTAGCTGAAAATGTTG 2146
 QY 1592 ATTTCCTATGCTAAAGATGGATCTACAAAAGNAGATCAATTAAGTTGATGCAATTAGCAG 1651

Db 2147 AAGAAAGTGTAGCTGAAAATGTTGAAGAAATCGTAGCTCCAACCTGTTGAAGAAATCGTAG 2206
 QY 1652 TAGGAGCTGTTGCTCGATATAAACTTGTCTCCAGTTGTA 1689
 Db 2207 CTCCTCACTGTTGAAGAAATGTTAGCTCCAAGTGTGTA 2244

RESULT 2
 US-08-973-462-1
 ; Sequence 1, Application US/08973462B
 ; Patent No. 6191270
 ; GENERAL INFORMATION:
 ; APPLICANT: DRUILHE, PIERRE
 ; APPLICANT: DAUBERSIES, PIERRE
 ; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
 ; FILE REFERENCE: 0650-0125-0 PCT
 ; CURRENT APPLICATION NUMBER: US/08/973,462B
 ; CURRENT FILING DATE: 1998-02-06
 ; EARLIER APPLICATION NUMBER: PCT/FR96/00894
 ; EARLIER FILING DATE: 1996-06-12
 ; EARLIER APPLICATION NUMBER: FR 95/07007
 ; EARLIER FILING DATE: 1995-06-13
 ; NUMBER OF SEQ ID NOS: 29
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 6152
 ; TYPE: DNA
 ; ORGANISM: P. falciparum
 ; US-08-973-462-1

Query Match 5.2%; Score 95.6; DB 3; Length 6152;
 Best Local Similarity 43.6%; Pred. No. 6.2e-11;
 Matches 645; Conservative 0; Mismatches 809; Indels 24; Gaps 4;

QY 233 CAAATGAGAAAGATTTCTTCACTCCAACCTGGAGATAAAAAAGTTTATTTCAGAACAAACTT 292
 Db 1014 CAAAGTGTGAAGAAAGTAGTCTTCAAGTGTGTGAAGAAAGTAGATTAAGTATTAAGTATGAAG 1073
 QY 293 TAACTACAGCTAATGAAATGAAGATTTATGAAAGACAA---CTTTAAAAAATTTAGATG 349
 Db 1074 AAAATGTAGCTCCAACCTGTTGAAGAAATCGTAGCTCCAAGTGTGTGAAGAAAGTGGCTC 1133
 QY 350 CAGGAGATATGCTATTATAGATTAACTTATAATAATGCTTAAACTGTTGAAATTAAG 409
 Db 1134 CAAAGTGTGAAGAAAGTAGTAGAGAAATGTTGAAGAAAGTAGCTGAAAATGTTGAAG 1193
 QY 410 TAGTAGCAGCTAGTCAAAAAACAGTAGTTGTATCTAGTATGCGAAAAAATAGTCAAAAG 469
 Db 1194 AAAGTGTAGCTGAAAATGTTGAAGAAAGTAGCTGAAAATGTTGAAGAAAGTAGCTG 1253
 QY 470 ATATAGCTGAAAAATPATGTTTGAAGACAAAGACTTAGAAAAATCGACTTAAAAAATATAA 529
 Db 1254 AAAATGTTGAAGAAATCGTAGCTCCAACCTGTTGAAGAAATCGTAGCTCCAACCTGTTGAAG 1313
 QY 530 ATGCTCTCAGATTTCAAGTAAAACTGATGATTACTATCAAGTAGTCTTTTATCCAAAGGAA 589
 Db 1314 AAAATGTAGCTCCAAGTGTGTGAAGAAAGTGGCTCCAAGTGTGTGAAGAAAGTAGAG 1373
 QY 590 AGAGATTACAAAGTTTCTCACTTATAGAGCTACAAAATTTAATAAGAAAGTGCATATG 649
 Db 1374 AAAATGTTGAAGAAAGTAGCTGAAAATGTTGAAGAAAGTAGCTGAAAATGTTGAAG 1433
 QY 650 GTAATACACAGTAATTAATTAATCTTAAAACTCTAGTAGTAAGAGTAAATTTAAAGACTGCAG 709
 Db 1434 AAAAGTGTAGCTGAAAATGTTGAAGAAAGTAGTAGCTGAAAATGTTGAAGAAAGTAGCTG 1493
 QY 710 TAGAAGAGTTACAAAATTTGAATGCTAGTTATTCTTAATCTACAACCTTTAGCTGGTGAG 769
 Db 1494 AAAATGTTGAAGAAATCGTAGCTCCAACCTGTTGAAGAAATCGTAGCTCCAACCTGTTGAAG 1553
 QY 770 ACAGAAATCAAAACAGCTATAGAGTAAGTAAGAAATATTAACAATATGATGCGCAAAAT 829

Db 1554 AAATTGTAGCTCAAGTGTGTAGAAAGTGTGGCTCCAAAGTGTGTGAAGAAAGTGTAGAAG 1613
QY 830 CAGATCAATTCAGCTGATGTTAAAGAGAAATGTTAAATGTTGTATTTAGTAGTGTGCAATG 889
Db 1614 AAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAG 1673
QY 890 CACTAGTAGATGATTTGGCTCCCTTTAGCAGCAGCAAGAAAGTGTCCACTATTAT 949
Db 1674 AAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTG 1733
QY 950 TAACTTTCAAAAGATATAATTTAGATTTCTGCTAGTAAA-----ATCTGAAATAAAGAGAG 1000
Db 1734 AAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAG 1793
QY 1001 TTTTAGACTTTAAAACTTCAAGAGATTAACAGGAAACAGTTTATATAGCTGGTGAG 1060
Db 1794 AAATCGTAGCTCAACTGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATGTTAGCTC 1853
QY 1061 TTAATAGTGTATCTAAAGAAAGTGTAAACAGAAATTAGAAATCAATGGGATTTAAAGTTGAAA 1120
Db 1854 CAAGTGTGTGAAGAAAGTGTGGCTCCAGTGTGAAGAAAGTGTAGAGAAATGTTGAAG 1913
QY 1121 GATTCTCAGGTGATGATGATGAAATCTTTTAAATAATGAGGTGAAATAGGCTTAG 1180
Db 1914 AAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTG 1973
QY 1181 ATAATGATAAGGCTTATGTAGTTGGTGGAAACAGATTTAGCAGATGCCATGAGTATAGCTT 1240
Db 1974 AAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAATCGTAGCTCCAACTGTTGAAG 2033
QY 1241 CAGTTCCTTCTACTAAATAGATGTTAATG-----GTGTTGTAGATGAGAAACAAATG 1291
Db 2034 AAATCGTAGCTCAACTGTTGAAGAAATTTAGCTCCAACTGTTGTGAAGAAAGTGTGGCTC 2093
QY 1292 GACATGCTACTCAATAGTGTGTAGATGGAAGAAAGCTGTAAATAATCTGTAGTACTTAG 1351
Db 2094 CAAGTGTGAAGAAAGTGTAGAGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAG 2153
QY 1352 ATAGTTCCTTGAAGAGCGCTGATGATGATATATAGTGGATTTGCAAGTGTATCTGAAA 1411
Db 2154 AAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAATCGTAGCTC 2213
QY 1412 AGATGAAGAGCTATATCAGATGCTACTGTGTAAAGCGGTTTACAAGAGTTTAAAGGCGAG 1471
Db 2214 CAACTGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATTTAGCTCCAACTGTTGTAG 2273
QY 1472 ATAGACAAGACACTAACTCTGAAGTTTATAAACAATATTTATGCTTAATGATCTGAATAG 1531
Db 2274 AAAGTGTGGCTCCAACTGTTGAAGAAAGTGTAGAGAAATGTTGAAGAAAGTGTAG 2330
QY 1532 CTAAGCTGCAGTTCCTTGAAGAAAGTGTAGTGAATTCAGGTGCTTCAAGTGTAGTGCAGGATTTTA 1591
Db 2331 CTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTG 2390
QY 1592 ATTTCTATGAGCTTAAAGATGATCTACAAAGAGATCAATTTAGTTGATGATCTAGCAG 1651
Db 2391 AAGAAAGTGTAGCTGAAATGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATCGTAG 2450
QY 1652 TAGCAGCTGTGGCTGGATATAAACTGTCTCCAGTTGTA 1689
Db 2451 CTCCAACTGTTGAAGAAATGTAGCTCCAACTGTTGTA 2488

RESULT 3

5231168-1

; Patent No. 5231168

; APPLICANT: DIEGIEL, MORTEN; BORRE, MARTIN; JEPSEN, SOREN;

; VUUST, JENS; RIENECK, KLAUS; WIND, ANNETTE; JAKOBSEN, PALLE H.

; TITLE OF INVENTION: MALARIA ANTIGEN

; NUMBER OF SEQUENCES: 19

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/409,658

; FILING DATE: 18-SEP-1989

; SEQ ID NO: 1:
; LENGTH: 3095

5231168-1

Query Match 4.8%; Score 87; DB 6; Length 3095;

Best Local Similarity 43.7%; Pred. No. 3.5e-09;

Matches 436; Conservative 0; Mismatches 555; Indels 1;

QY 165 AGCAGCACAAAACTTTAAACACAGGTGCGAGTTATTTTGAACAAAGATACAAAAAGTTTACTTTT 224
Db 1194 AGTTGAACATGAATAGTAGAGTTGAAGAAATTTCTACCAGAGATATAAAATGAAAAAGG 1253
QY 225 CTATGATGCAATGAGAAAGATTTCTCAACTCAACTGCGAGATATAAAAGATTATTCAGA 284
Db 1254 TCAACATGAAATAGTAGAGTTGAAGAAATTTCTACCAGAGATGATATAAAATGAAAAAGT 1313
QY 285 ACAAACTTTTAACCTACAGCTTAATGGAATGAAGATTATGTAAAGACAACTTTAAAAAATTT 344
Db 1314 TGAACATGAAATAGTAGAGTTGAAGAAATTTCTACCAGAGATATAAAATGAAAAAGTCA 1373
QY 345 AGATGCAAGAGATATGCTTATTTATAGATTTAACTTTATATAATGCTTAAAACTGTTGAAAT 404
Db 1374 ACATGAAATAGTAGAGTTGAAGAAATTTCTACCAGAGATATAAAATGAAAAAGTTGAACA 1433
QY 405 TAAAGTAGTAGCAGCTAGTGTGAAAAACAGTAGTTGTCTTAGTGATGCCGAAAAATAGTGC 464
Db 1434 TGAATAGTAGAGTTGAAGAAATTTCTACCAGAGATATAAAATGAAAAAGGTCAACATGA 1493
QY 465 AAAAGATATAGCTGAAAAATATGTGTTGAAGACAAAGACTTAGAAAAATGCCTAAAAAC 524
Db 1494 AATAGTAGAGTTGAAGAAATTTCTACCAGAGATATAAAATGAAAAAGTTCACATGAAAT 1553
QY 525 TATAATGCTCAGATTTTCAGTAAAACTGATAGTTACTATCAAGTAGTTCTTTTATCCAAA 584
Db 1554 AGTAGAAGTTGAAGAAATTTCTACCAGAGATATAAAATGAAAAAGGTCAACATGAAATAGT 1613
QY 585 AGGAAAGAGATTACAAGTTTCTCAACTTATAGAGCTACAAATTTATAATGAAAGAACTGC 644
Db 1614 AGAGTTTGAAGAAATTTCTACCAGAGATATAAAATGAAAAAGGTCAACATGAAATAGT 1673
QY 645 ATATGTTAATACACAGTAAATTTAACTCTAAATCTACTAGTAAGAGTAATTTTAAAGAC 704
Db 1674 AGAGTTTGAAGAAATTTCTACCAGAGATATAAAATGAAAAAGGTCAACATGAAATAGT 1733
QY 705 TGCAGTAGAGAGTTTACAAAAATTTGAATGCTAGTTATTTAATACTACAACTTTAGCTGG 764
Db 1734 GGTGAAGAAATTTCTACCAGAGATATAAAATGAAAAAGTTCAACATGAAATAGTAGAGT 1793
QY 765 TGAATGACAGAAATACAAACAGCTATAGAGATAAGTAAAGAAATATTACAATTAATGATGCCGA 824
Db 1794 TGAAGAAATTTCTCCAGAAATTTGTTGAAATTTGAAGAAAGTACCATTACAAACAAATAACAA 1853
QY 825 GAAATCAGATCATCTAGCTGATGTTAAAGAGAAATGTTAAAAATGTTGTATTAGTAGTGC 884
Db 1854 T-----GAAAAATTTGAAACTATAAAACCCAGAGAAATAAAGAAATGAAATTTAGTGTGA 1907
QY 885 AAATGCACCTAGTAGATGATTTAGTTCGGCTCTTTAGCAGCAGAGAAAAAGATGCTCCACT 944
Db 1908 AGAAAAAGCAATTTCCACAGAACCCGCTGCTACCTACATTAATGAAATGAAACCTGATAC 1967
QY 945 ATTATTTAACTTTCAAAAGATAAATTTAGATTTCGTGATAAAACTGAAAAATAAGAGATTTT 1004
Db 1968 TCCCAACCACTCTGAAGGTGAATCCACTAAACCCAGATATATGTTCAAAATTTAAATAGTACA 2027
QY 1005 AGACTTAAAACTTTCAACAGAGATAACAGAGAAAAACAGTTTATATAGCTGGTGGAGTTAA 1064
Db 2028 AGAAAAATAAACCAAAATAAAAGGAAAAACACAGTAGTAGATGGTCCAAAAACATGTAGAAC 2087
QY 1065 TAGTGATCTTAAAGAGTTCTTAACAGAAATTTAGAAATCAATGGGATTTAAAGATTGAAAGATT 1124
Db 2088 AAATATACAGAGATGATTAATGATGAAGAGGATGATGATGATATAGATTTTGAAGATT 2147
QY 1125 CTCAGGTGATGATAGATATGAAACTTCTTTAAAAATA 1161

Db 5084 CAGTAGTCTGCAACGCTGATATAGATAATCTGTCAGCAAAATGATGTGGATAATG 5143
Qy 272 AGTTTATTCAGAAACAACTTTAACTACAGCTTAATGGAATGAGATTTATGTAAGACAA 331
Db 5144 CAAAACCTACAAATGAAGCTTACAACTCGCAGCCATTTACCTGATGCAAAATGTTAAACCG 5203
Qy 332 CTTTAAAAAATTTAGATGAGGAGAAATATGCTATTATAGATTTAACTTATTAATTAATGCTTA 391
Db 5204 CAGCAAAAACAGCAATTTGAGATAAAGTACAGCTCAAGAAACAGCAATTTGATGGAAATA 5263
Qy 392 AAACCTGTTGAAATTTAAAGTAGTAGCAGCTAGTGAAGAAACAGTAGTTGTTATCTAGTGATG 451
Db 5264 ACGGCTCAACAACTGAAGAAAGAGCAGCTGCTAAACAAACAGTTTCAAACTGAAGAAACAA 5323
Qy 452 CGAAAAATAGTGCAAAAGATATAGCTGAAAGAAATATGTTTGAAGCAAAAGCTTAGAAA 511
Db 5324 CAGCTGATGCGCAATAGATGAGCAGCATACAAATGCGGAAGTTGAAGCGGCTTAAAGAAAG 5383
Qy 512 ATGCACCTAAAACTATAAATGCCTCAGATTTTCAGTAAAACTGATAGTTACTATC---AAG 568
Db 5384 CAGCAATTTGCTAAATTTGAAGCAATTCAGCCAGCAACAACTAAAGATAATGCGAAG 5443
Qy 569 TAGTCTCTTATCCAAAGGAAAGAGATTAACAGGTTTCTCAACTTATAGAGCTTACAAAT 628
Db 5444 AAGCAATTTGCTACGAAGCGAATGAACGTTAAACAGCAATCGCTCAAAACGCAAGACATTA 5503
Qy 629 ATAAAGAGAACTGCAATATGTTAATACAGCAGTAAATTAATCTTAAATCTTACTAGTA 688
Db 5504 CTGCTGAAGAAATTTGCAAGCGGCTTAATGCGGACGTAGATATGCTGTGACACAAAGCAATA 5563
Qy 689 AGATGAATTTAAAGACTGC-----AGTAGAAGAGTTTACAAAATTTGAATGCTAGTT 739
Db 5564 GCACATTTGAAGCTGCTTAATAGTCAAAATGATGTAGCAAGCGAAACGACAGGTGAAA 5623
Qy 740 ATTCTAATCTACAACTTTAGCTGGTGATGACAGAAATACAAACAGCTTATAGAGATAAGTA 799
Db 5624 ATAGTATTGATCAAGTAACCAACCAACAGTTAATAAAAAAGCACTGCACGCTAATGAAATCA 5683
Qy 800 AAGAAATTAACAAATATGATGGCGAGAAATCAGATCATTCAGCTGATGTTTAAAGAGATG 859
Db 5684 CAGCAATTTTAAATAACAAATTTGCAAGAGATTTCAAGCTACGCCAGATGCAACAGATGAAG 5743
Qy 860 TTAAGATGTTGTTATAGTAGTAGTGCAATGCACCTAGTAGATGATGATTAGTTGCGGCTCCT 919
Db 5744 AAAAAACAGCAGCTGATGCTGAAGCAATTAATGAAATGTTAAAGCAATCAAGCCATTT 5803
Qy 920 TAGCAGCAGAAAAAGATGCTCCACTATTATTAATCTCAAAAGATAAATTAGATTGCTGCAG 979
Db 5804 CAGCAGCACTACTAACGCACAGTTGATGAAGCT---AAGCAAAATGCAGAAAGCAGCA 5860
Qy 980 TAAATCTGAAATTAAGAGAGTTTATAGACTTTAAACTTCAACGAAGTAAACAGGAAAAA 1039
Db 5861 TTAATGCGGTAAACCAAAAAAGTTGTAAGAAACAAAGCGGCTTAAAGATGAAATTTGATCAAT 5920
Qy 1040 CAGTTTATAGCTGGTGGAGTTAATAGTGTATCTTAAGAGTTGTTAAACAGAAATTAAGAT 1099
Db 5921 TACAAGCAACGCAAAACAAATGTTATCAATATGATCAGACGCTTACAAACAGAAAGAAAG 5980
Qy 1100 CAATGGGATTTAAAAAGTTGAAAGATTTCTCAGTGATGATAGATATGAAGCTTTCTTTAAAAA 1159
Db 5981 AAGCAGCTATTCAACATTTAGCAACAGCAGTTACAGACGCGGAAAAATATATTACAGCTG 6040
Qy 1160 TAGCAGGTGAAATAGGCTTATAGATAATGATAGAGCTTATGATGTTGGGAGCAGGATTAG 1219
Db 6041 CAACCTGATGATAATGGTGTAGATCAGGCGAAAGACGCTGGAAGAAATTTCAATTTCAAGCA 6100
Qy 1220 CAGATGCCATGATAGCTTACGTTGCTTCTACTTAAATTAGATGTTAAATGGTGTGCTGTAG 1279
Db 6101 CGCAACAGCAACAGCGGTTAAATCAATGCTTAAATATGATGTTGATCAAGCTGTGACAA 6160
Qy 1280 ATAGCAACAAATGGACATGCTACTCTCAATAGTTGTTGATAGTGAAGAAAGCTGATAAATAT 1339

Db 6161 CTCAAAATCAAGCAATTTGATTAATACAACTGGTGCTCAACTGAAGAAAAAATGACGAA 6220
Qy 1340 CTGATCACTTAGATAGTTTCTTTAGGAAGCGCTGATGTAGATATAATPAGGTGATTTGCAA 1399
Db 6221 AAGATTAGTTTAAAGCTTAAAGAAAGCGTATCAAGATATCTTAAATGCAAAACAA 6280
Qy 1400 GTGTATCTGAAAGATGGAAGAGCTATATCAGATGCTACTGTTAAAGCGGTTACAAGAG 1459
Db 6281 CTAATGATGTTTACGCAAAATTAAGAGATCAAGCAGTTGCTGATATTTCAAGGTATTTACTGCAG 6340
Qy 1460 TTAAGGCGACGATAGCAAGACACTTAACCTCTGAAGTTTATAAAAAATATTTATGCTAAATG 1519
Db 6341 ATACAAAC---AATTAAGATGTTGGAAGAGATTAATAGCAACAAAGCAACGAAACAA 6397
Qy 1520 ATACTGAAATAGCTAAAGCTGCAGTTTTTATAGATAAAGATTCAGGTGCTTCAAGTAGTGATG 1579
Db 6398 AAGCGCTTATTGCAAACTGCGACTACTGAGCAAAAGAAAGCAAGCAAAATCAAC 6457
Qy 1580 CAGGAGTATTTAATTTCTATGATGCTAAAGATGGATCTACAAAGAGATCAATTTAGTTG 1639
Db 6458 AAGTAGACGCACAAATTTAAACAAGGTAATCAAAATATTTGAAATATGCAAGTCAATCGATG 6517
Qy 1640 ATGCATTAGCAGTAGGAGCTGTTGCTGGATATAAACTTCCTCCAGTTGTTATTTAGCTACTG 1699
Db 6518 ATGTAACACTGCAAAAGATTAATGCAATTCAGCAATTCACCAATTTCAAGCATCAACAG 6577
Qy 1700 ATTCTTTATCTCTGATCAATCGGTTGCTTATAGCAAAAGTTTGTAGAGAAAAATATTTCTA 1759
Db 6578 ATGTTAAACGATGCAAGCGGGAATTCCTAACTGAATGCAA--AATAAATAACTGA 6635
Qy 1760 AAGATTTAAACAAAGTTGCTCAAGGAAT 1787
Db 6636 AATACTTAATAATAGACTACTAAT 6663

RESULT 6

US-08-781-986A-63
; Sequence 63, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8155 base pairs
; TYPE: nucleic acid

```

; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-63

```

Query Match 4.7%; Score 85.2; DB 4; Length 8155;
Best Local Similarity 43.1%; Pred. No. 1.1e-08;
Matches 685; Conservative 0; Mismatches 883; Indels 20; Gaps 5;

Db	6101	CGCAACAGCGAACAGCGGTTAAATCAAAATGCTTAAAAATGATGTTGATCAAGCTGTGCACAA	6160
Qy	1280	ATAGAACAATGGCAGTCCTACTCCAAATAGTGTGTGTAGATGGAAAGCTGTGATAAAATAT	1339
Db	6161	CTCAAAATCAAGCAATTTGATAATACAACCTGGTGTCTACAACTGAAGAGAGAAAAATGCGACAA	6220
Qy	1340	CTGATGACTTAGATAGTGTTCCTTAGGAGCGCTGATGTAGATATATAATAGGTGGATTTGGCAA	1399
Db	6221	AAGATTTAGTGTTTTAAAGCTTAAAGAAAGCGTATCAAGATATCTTAAATGCGACAAACAA	6280
Qy	1400	GTGTATCTGAAAGAGATGGAAGAGCTATATCAGATGCTACTGTGTAAGGCGTTTACAAGAG	1459
Db	6281	CTAATGATGTTACGCAAAATTAAGNATCAAGCAGTGTCTGATATTTCAAGGTATTACTGCAG	6340
Qy	1460	TTAAGCGGAGATAGACAAGACACATTAACCTCTGAAGTTATAAAAAATATTTAGCTTAATG	1519
Db	6341	ATACAAC---AATTTAAAGATGTTTCGAAAGATGAAATTAGCAACAAAAGCAACGAAACAA	6397
Qy	1520	ATACTGAAATAGCTTAAAGCTGCAGTGTTTAGATAAAGATTCAGGTGCTTCAAAGTAGTGAATG	1579
Db	6398	AAGCGCTTTATTGCAACAACTGCGAGATCGGACTACTGAAGAAAAAGAACAGCAATCAAC	6457
Qy	1580	CAGGAGTATTTAAATTTCTATGTAGCTAAAGATGGATCTACAAAAGAGATCAATATTAGTTG	1639
Db	6458	AAGTAGCGCCAAATTAACACAAGGTAATCAAAAATATTGAAAAATTGCAAGTCAATCGATG	6517
Qy	1640	ATGCATTTAGCAGTAGGAGCTGTTGCTGGATATAAACTTGCTCCAGTTGTATTAGCTACTG	1699
Db	6518	ATGTAAACACTGCAAAAAGATAATGCAATTCAGGCAATTTGACCCCAATTTCAAGCATCAACAG	6577
Qy	1700	ATTCCTTTATCTTCGTGATCAATCGGTTGCTATAAGCAAAAGTTGTAGGAGAAAAATATTCTA	1759
Db	6578	ATGTTAAACGAATGCAAGGCGAAATGCTTAATCGAATGCAA--AATAAATAACTGA	6635
Qy	1760	AAGATTTAAACAAGATTGGTCAAGGAAT	1787
Db	6636	AATACTTAATAATATGAGACTACTAAT	6663

RESULT 7
US-09-754-947-2
; Sequence 2, Application US/09754947
; Patent No. 6828110
; GENERAL INFORMATION:
; APPLICANT: Lee, Bruce Andrew
; APPLICANT: Flores, Becky Mar
; APPLICANT: Walkers, Gunars Edwin
; APPLICANT: Biosite Diagnostics, Inc.
; TITLE OF INVENTION: Assays for Detection of Bacillus Anthracis
; FILE REFERENCE: 014907-003310US
; CURRENT APPLICATION NUMBER: US/09/754,947
; CURRENT FILING DATE: 2001-01-04
; PRIOR APPLICATION NUMBER: US 60/174,901
; PRIOR FILING DATE: 2000-01-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2370
; TYPE: DNA
; ORGANISM: Bacillus anthracis
; FEATURE:
; OTHER INFORMATION: surface array protein (SAP)
US-09-754-947-2

Query Match 4.5%; Score 82.2; DB 4; Length 2370;
Best Local Similarity 42.3%; Pred. No. 3.5e-08;
Matches 715; Conservative 0; Mismatches 958; Indels 18; Gaps 4;

QY	1100	CAATGGGGTAAAAAGTTGAAAGATCTCAGGTGATGATAGAGAATGAAACTTCTTAAAAA	1115
QY	1100	CAATGGGGTAAAAAGTTGAAAGATCTCAGGTGATGATAGAGAATGAAACTTCTTAAAAA	1115
Db	5981	AAGCAGCTATTCAACAATTACCAACAGCAGTTTCAGACGGGAAAAATAATTATTACAGCTG	6040
QY	1160	TAGCAGGTGAAATAGGCTTAGTAATGATTAAGGCTTAATGTAGTTGGTGGAAACAGGATTAG	1219

QY 2 TGAATAAAGAAATTTAGCAATGGCTATGGCAGCTGTCTTACTGTAGTGTCTTCTGTGCTC 61
 Db 668 TTAAGAGGTAACCTTTTACAGAGATAAAGACTCTCTACAGTTGAATTTATATAGTAACT 727
 QY 62 CAGTTTTTGCAGCAGCTTCAGATGTAAATTCACCTAACAAGTGTACAAATGATAGTATA 121
 Db 728 TAGCAGCTAAACAACTTACACTGTAGATGTAAACAAAGTTGGTAAACAGAGTAGCTG 787
 QY 122 CAGTATCAAAATCTAAAGCTAGTACTTGTAGTAAGGATATTTTACGAGCACAACCTTAA 181
 Db 788 TAGGTTCTTTAGAGCAAAAACATCGAAATGGCTGACCAACAGTTGTAGCTGATGAGC 847
 QY 182 CACAGGTGCAGTTATTTTGAACAAAGATACAAAGTTTACTTTCTATGATGCAAAATGAGA 241
 Db 848 CACAGCAATTTACAACTTACAGTTTAAAGATGAAACGGTACTGAAAGTTGTTTCCACGAGG 907
 QY 242 AAGATCTTCAACTCACTGAGATATAAAGTTTATTCAGAACAACTTTTAACTTACAG 301
 Db 908 GTATTGAATTTTGAACGCGCAGCTGCAGAAAATTAATGCAAAAGGTGAAATCACCTTTAG 967
 QY 302 CTAATGGAATGAAGATTATGTAAAGCAACTTTTAAATAATTTAGATGCGAGGAGATATG 361
 Db 968 CAAAGGTACTTCAACTACTGTCTGAAAGCTGTTTATAAAGAGCGGTAAAGTAGTAGCTG 1027
 QY 362 CTATTATAGATTTAACTTATAATATCTTAAACTCTGTGAAATTTAAAGTAGTAGCAGCTA 421
 Db 1028 AAGTTAAGAGTA-----AAGTTTCTCTGAGGTGCTGAGTAGCTTCAATCT 1078
 QY 422 GTGAAAAAAGCAGTAGTTGTATCTAGTGTGCGAAAAATATGTGCAAAAGATATAGCTGAAA 481
 Db 1079 CTAACTGGACAGTTGCGAAGCAAAATAAAGCTGACTTTTACTTCAAGAGATTTTCAACAAA 1138
 QY 482 AATATGTTTGAAGCAAGACTTAGMAAATGCACTTAAACTATATAATGCTCAGATT 541
 Db 1139 ACAATAAGTTTACGAAGGCGCAACGCTTACGTTCAAGTAGAATTTGAAAGATCAATTTA 1198
 QY 542 TCAGTAAACCTGATAGTTACTATCAAGTAGTCTTCTTTATCCAAAGGAAAGAGATTACAAG 601
 Db 1199 ACCGAGTACCACTGGAAGTTGAAATATGATGCTGTTTAAACACAGAGTTGCTGTAGTAG 1258
 QY 602 GTTCTCAACTTATAGAGCTACAAATTTATATGAAGAACTGCAATATGTTAAATACACCAG 661
 Db 1259 ATAAAGCTACTGGTAAAGTAAGTATCTCTGAGGAAAGCACCAGTAAAGTAAGTAACTG 1318
 QY 662 TAATA---TTAACTTAAATCTACTAGTAAAGATTAATTAAGACTGCGAGTAGAGAGT 718
 Db 1319 TAAAGATTCAAAAGTAAAGCACTTGTGTTTCAACACAGTTGAATTTGAAGCTTTTCGCTC 1378
 QY 719 TACAAAATTTGAATGCTAGTTATTTCTAATACTCAACTTTAGCTGGTGTAGTACAGAGATAC 778
 Db 1379 AAAAGCATGAAGACATTTAAATTAGAAAATACTAACGTAGTGGCTTTCTACAAAAGATG 1438
 QY 779 AAAAGCATGATAGATAAGTAAAGATTTTACAAATATGATGCGGAGAAATCAGATCAAT 838
 Db 1439 TAACAGATTTAAAGTAAAGGCTCCAGTACTAGTCAATCAACGCTAAAGAGTTTACAGCTC 1498
 QY 839 CAGCTGATTTAAAGAGAACTTTAAATAATGTTTATAGTAGTGGTCAATGCTAGTAG 898
 Db 1499 CTGTAACAGTGAAGTACTTGTATTAAGATGTTAAAGAAATTTAAAGAACAAAATTTAGAAG 1558
 QY 899 ATGATTTAGTTGGGCTCTTTAGCAGCAAAAAGATGCTCCACTATTATTAACTTCA 958
 Db 1559 CTAATATGTGAACAGAGAAATTTAGTTCTGATGCGAGGTCAAGAGCTGGTAAATTTATA 1618
 QY 959 AAGATAAATTTAGTTCTGATGAAATTTCTGAAATTAAGAGAGTTTGTAGACTTTAAACCTT 1018
 Db 1619 CAGTTGTTTAACTGCAAAATCTGGTGAAGAAAGCAAAAGCTTACATTAGCTCTAGAAAT 1678
 QY 1019 CAACAGAGTAAACAGGAAAAAGCTTTTATATAGCTGGTGGAGTTTATAGTGTATCTAAAG 1078
 Db 1679 TAAAGCTCCAGGTGCACTTCTTAAATTTGAAGTTGCTGTTTGTAGACACAGAAATTAGATA 1738
 QY 1079 AAGTTGTGAACAGAAATTAAGATCAATGGATTAAAGTTTGAAGATTTCTCAGGTGATGATA 1138

Db 1739 AATATGTTTACTGAGGAAACCAAAAGAAATGCAATGACTGTGTTTCAGTTCTTCTGTAGATG 1798
 QY 1139 GATATGAACCTTCTTTTAAATAATAGCAGGTGAAATAGGCTTAGATTAATGATAAGGCTTATG 1198
 Db 1799 CAAATGGATTTAGTATTTAAAGAGTGCAGAAAGCAGCTGAATTAAGTAAACACACAAACA 1858
 QY 1199 TAGTTGGTGGAAACAGGATTTAGCAGATGCCATGATAGCTTTCAGTTGCTTCTTACTAAAT 1258
 Db 1859 AAGAAGGT---AAGAAGTAGACGCAACTGATGCGCAACAGTTACTGTACAAAATAACAGTG 1915
 QY 1259 TAGATGGTAAATGCTGTTGTAGATAGAAACAAATGGACATCTACTCCAATAGTTGTTAGTAG 1318
 Db 1916 TAAATTTACTGTTGGTCAAGGTGCAAAAGCTGGTGAGACTTTATAAGTAAACAGTTGTACTAG 1975
 QY 1319 ATGGAAGAGCTGATAAAATATCTGATGACTTATAGATAGTTTCTTAGGAAGCGCTGATGTAG 1378
 Db 1976 ATGGTAAATTAATCAACTCACTCAATTCATTAAGTTGTTGATACAGCACCACTGCTAAAG 2035
 QY 1379 ATATAATAGGTGGATTTGCAAGTGTATCTGAAAAGATGGAAGAGCTATATCAGATGCTA 1438
 Db 2036 GATTAGCAGTAGAATTTTACAAGCACATCTCTTAAAGAGTAGCTCCAAATGCTGATTTAA 2095
 QY 1439 CTGGTAAAGGCGTTTCAAGAGTTAAAGGCGACATAGACACACACTAACTCTGAAAGTTA 1498
 Db 2096 AAGCTGCACCTTTTAAATATCTTATCTGTTGATGGTGTACTGCGACTACAGCAAAAGCAA 2155
 QY 1499 TAAACATATTTATCTAATGATGATGCTGAAATAGTAAAGCTGAGTTTGTAGATAAAGATT 1558
 Db 2156 CAGCTTCTAATGTAGAAATTTGTTCTGCTGACACAAATGTTGTAGCT---GAAATGGTA 2212
 QY 1559 CAGGTGCTTCAAGTAGTAGTATGATGAGGAGTATTTAAATTTCTATGTAGCTAAAGATGATCTA 1618
 Db 2213 CAGTTGGTCAAAAGGTGCAACATCTATCTATGTGAAAACCTGACAGTTGTAAGAGATG 2272
 QY 1619 CAAAAGAGATCAATTAGTTGATGCAATTAGCAGTAGGAGCTGTTGCTGGATATAAAGCTTG 1678
 Db 2273 GAAAAGAGCAAAAGTAGAATTTGTATAAGCTGTACAGTTGCAAGTTTCTTATTTAAAGAAG 2332
 QY 1679 CTCAGTTGTA 1689
 Db 2333 CAAAACCTGCA 2343

RESULT 8

US-09-902-540-1280
 ; Sequence 1280, Application US/09902540
 ; Patent No. 6833447
 ; GENERAL INFORMATION:

; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Wiegand, Roger C.
 ; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
 ; FILE REFERENCE: 38-10(15849)B
 ; CURRENT APPLICATION NUMBER: US/09/902,540
 ; CURRENT FILING DATE: 2001-07-10
 ; PRIOR APPLICATION NUMBER: 60/217,883
 ; PRIOR FILING DATE: 2000-07-10
 ; NUMBER OF SEQ ID NOS: 16825

; LENGTH: 1039

; TYPE: DNA

; ORGANISM: Myxococcus xanthus

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)..(1039)

; OTHER INFORMATION: unsure at all n locations

US-09-902-540-1280

Query Match 4.2%; Score 76; DB 4; Length 1039;
 Best Local Similarity 44.2%; Pred. No. 5.8e-07;
 Matches 408; Conservative 0; Mismatches 507; Indels 9; Gaps 2;

```
QY 267 TAAAAAGTTTATTGAGACCAACTTTAACTACAGCTAATGGAATGAGATTATGTAAA 326
Db 69 TAAAAACCAACCAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 128
QY 327 GACAACTTTAAAAATTTAGATGCGAGGAAATATGCTATTATAGATTAACTTATAATAA 386
Db 129 AAAATAAAAACAATATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 188
QY 387 TGTCTAAACTGTTGAAATTTAAAGTAGTAGCAGCTAGTGAAAAACAGTAGTTGTATCTAG 446
Db 189 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 248
QY 447 TGATGCGAAAAATAGTGCAGAAAGATATAGCTGAAATAATATGTTTGAAGACAAAGACTT 506
Db 249 AAAAAACAATACATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 308
QY 507 AGAAATGCACTAAAAACTATAAATGCGCTCAGATTTCAGTAAACTGATAGTTACTATCA 566
Db 309 ATAAAAAAA--AAAAATTAATATCAATCATATAAATAAAAAATAAAAAATAAAAAATA 366
QY 567 AGTAGTCTTTATCCAAAGGAAGAGATTACAAGTTTCTCAACTTATAGAGCTACAAA 626
Db 367 AAAAAACCAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 426
QY 627 TTATATGAGGNACTGTCATATGTTAATACACAGCTAATATTAACCTTAAATCTACTAG 686
Db 427 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 486
QY 687 TAAGAGTAAATTTAAAGACTGCGAGTAGAGAGTTTACAAAAATTTGAATGCTAGTTATTCTAA 746
Db 487 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 546
QY 747 TACTACAACTTTAGCTGTTGATGACAGATAACAAACAGCTATAGAGATAAGTAAGAATA 806
Db 547 TAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 606
QY 807 TTACATATATGTCGAGCAATCAGATCATTCAGCTGATGTTAAGAGAAATGTTAAAAA 866
Db 607 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 666
QY 867 TGTGTTATTAGTAGTGCATAATGCTAGTAGATGATAGTTGCGGCTCCTTTTAGCAGC 926
Db 667 TGTGTTTT-----CTTAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 719
QY 927 AGAAAGAGTGTCCATATTATTAATCTTCAAAAGATAAATTAGATTTCTCAGTAAATC 986
Db 720 AAGAAATTAACAAAGCTCTAATCAACACCCCAATATAACAAAAATAATCTTAAAAACAAA 779
QY 987 TGAATAAAGAGAGTTTATAGACTTAAACCTTCAACAGAGTAAACAGGAAAAACAGTTTA 1046
Db 780 AAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 839
QY 1047 TATAGCTGTGAGTTAATAGTGTATCTAAAGAGTTGTAACAGAAATTAGAATCAATGG 1106
Db 840 AGGGAATAAAAAACAATAAATACTTAAAAATAAAAAATAAAAAATAAAAAATAAAAA 899
QY 1107 ATTAAAAAGTTGAAGATTTCTCAGGTGATGATAGATAGAACTTCTTTAAAAAATAGCAGG 1166
Db 900 AATAAAAGAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 959
QY 1167 TGAATAGGCTTAGATATGATAA 1190
Db 960 AAAAAAATAAAAAATAAAAAATAA 983
```

RESULT 9

PCT-US93-07261-10

; Sequence 10, Application PC/TUS9307261

; GENERAL INFORMATION:

; TITLE OF INVENTION: PFBP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

```
; ADDRESSEE: John H. C. Blasdale
; STREET: One Giralda Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07261
; FILING DATE: 19930805
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,531
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blasdale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0288K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4766 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; STRAIN: Malayan Camp
; IMMEDIATE SOURCE:
; CLONE: p2b1;p12-1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..4766
; PCT-US93-07261-10
```

```
Query Match 4.0%; Score 73.4; DB 5; Length 4766;
Best Local Similarity 42.4%; Pred. No. 3.2e-06;
Matches 651; Conservative 0; Mismatches 871; Indels 15; Gaps 4;
```

```
QY 37 GTTACTGTAGTAGTTCTGCTGCCAGTCTTTTGCGAGCAGCTTCAGATGTATATCATA 96
Db 2124 GGTAAAGGAAGTAACCTAATAAAGCAGCGCAAAATCTAGAAGAATAATAATGAACTGATT 2183
QY 97 CAAGATGCTACAAATGATAAGTATACAGTATCAATATCTAAAGCTAGTGTAGTAAAG 156
Db 2184 GCAAAAGGTAAGGAAGTAACCTAATAAAGCAGCGCAAAATCTAGAAGAATAATAATGAACT 2243
QY 157 GATATTTTTCAGCAGCAAAAACCTTAAACAAGCTGCAGTTATTTTGAACAAAGATACAAA 216
Db 2244 GATTTAGCAAAAGGTAAGGAAGTAACCTAATAAAGCAGCAAAATCTGGAAGAATAATA 2303
QY 217 GTTACTTTCTATGATGCAAAATGAGAAAGATTCTTCAACTCCAACTCGGAGATAAAAAAGTT 276
Db 2304 GAAACCGATTTCAGCAAAAGGTAAGGAAGTAACCTAATAAAGCGCATGAAAACTAGAGAA 2363
QY 277 TATTCAGAACAACTTTTAACTACAGCTAATGGAAATGAAGATTATGTAAGACACTTTA 336
Db 2364 TATAATGAACCTGATTTTAGCAAAAGGTAAGGAAGTAACCTAATAAAGCAGCATGAAAACT 2423
QY 337 AAAAATTTAGATGCGAGGAAATATGCTA---TTATAGATTTTAACTTATAATAATGCTAAA 393
Db 2424 GAAGATATATGAATGAATGATTTAGCAAAAGGTAAGGAAGTAACCTAATAAGCACCGCA 2483
QY 394 ACTGTTGAAATTTAAAGTAGTAGTAGCTAGTGAAGAAAAACAGTAGTTGTATCTAGTAGTGG 453
Db 2484 AATCTAGAAGAATAATAATGAACCTGATTTAGCAAAAGGTAAGGAAGTAACCTAATAAGCA 2543
```



```

QY 715 GAGTTACAAAATTGATGCTAGTATTCTTAATCTACCACTTTAGCTGGTGATGACAGA 774
Db 11987 GTTATTATTCACAAAATTTAGCAATAATGACCCCACTAAACCTTAATATATACACA 11928
QY 775 ATACAAAACAGCTATAGAGATAAGTAAAGAAATATTACAATAATGATGGCGAGAAATCAGAT 834
Db 11927 CTTAAAAAGTGGTTTTAAAAACAATAATCAACCAATATTGATCTAATTAGTGAAGAG 11868
QY 835 CATTCAAGCTGATGTTAAAGAGAAATGTTAAAAATGTTGTTATTAGTAGGTCAAATGCACATA 894
Db 11867 CAATTAAGTGGTGATAATCATATTTTCATTTAAAAAACCAACAATAACTGCTAAAACTAAA 11808
QY 895 GTAGATGGATTAGTTCGGCTCTTTAGCAGCAGAGAAAAGATGCTCCACTATTATTAACT 954
Db 11807 GAAATATGATGATTATGAAATTTAGTTTTTCA-----AATCCTAGTTTTCAGCAAT 11760
QY 955 TCAAAAGATAAATTAGATTGCTCGTAAATCTGAAATTAAGAGAGAGTTTTAGACTTAAAA 1014
Db 11759 AAAAAAATTAAATTAACATTTTAAACCGGATGATTAATACTAATAACAAAAACGGTTGAA 11700
QY 1015 ACTTCAACAGAGTAACAGCAAAAAACAGTTTATATAGCTGGTGGAGTTAATAGTGTATCT 1074
Db 11699 GCTAGTATTGGTTAGATGGAAGAGCTATTTTAAAAACAAGTGATGATCAATTTTTCGA 11640
QY 1075 AAGAAGTTGTAACAGAAATAGAAATCAATGGATTAAGAGTTGAAGATCTCAGGTGAT 1134
Db 11639 CCAGATCATAAATATATACACTCAAAAAATGGAAGCTGATTAATAAAAAAGTTCTCTAATAT 11580
QY 1135 GATAGATATGAACCTCTTTAAAAATAGCAGGTGAAATAGGCTTAGAATAATGATAGGCT 1194
Db 11579 GATGAATTTCCATTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 11520
QY 1195 TATGATGTTGGTGAACAGATTTAGCAGATGCGATGATGATGATGATGATGATGATGATGAT 1254
Db 11519 GATAATAAGCATGAGTTTAAATATCCAGATCAAAAAATAAAGATTTAACTGCTGTTTAT 11460
QY 1255 AAATTAGATGTTAAGTGGTGTGTAGATAGAAATGATGATGATGATGATGATGATGATGAT 1314
Db 11459 AAGATATAAATAAACAATGAAATTCATGTTTCCAAATCAAAACTGATGATAAGGGAAGTT 11400
QY 1315 GTAGATGAAAAGCTGATAAATAATCTGAT 1344
Db 11399 ATGTTAATCCAAACATTAATTATTGAT 11370

```

```

RESULT 11
US-08-446-137B-1
; Sequence 1, Application US/08446137B
; Patent No. 6162903
; GENERAL INFORMATION:
; APPLICANT: Trowern, Angus R.
; APPLICANT: Atkinson, Anthony
; APPLICANT: Murphy, Jonathan P.
; APPLICANT: Laurence, Oliver S.
; APPLICANT: Duggleby, Clive J.
; TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED
; FROM L PROTEIN AND THEIR USES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,137B

```

```

Query Match 3.8%; Score 70; DB 3; Length 3279;
Best Local Similarity 42.8%; Pred. No. 1.5e-05;
Matches 468; Conservative 0; Mismatches 620; Indels 6; Gaps 2;

QY 99 AGATGGTACAAATGATAAGTATACAGTATCAATTAAGCTAGTACTAGTAAAGGA 158
Db 1374 AAAAGAAACACCCAGAAACACCCAGAAAGAACCCAGAAAGTACAAATCAAAAGTTAACTT 1433
QY 159 TATTTTAGCAGCACAAAACCTTAAACAAAGGTGTCAGTTATTTTGAACAAAGATACAAAAGT 218
Db 1434 AATCTTTCGAGATGGAAGATACAAACAGCAGAAATTCAAAGAAACATTTGAAGAGACAC 1493
QY 219 TACTTTCTATGATGCAAAATGAGAAAGATTCCTCAACTCCAACTGGAGATATAAAAAAGTTTA 278
Db 1494 AGCAAAAGCTTATGCTTATGCAAACTTATAGCAAAAGAAATGCGAATATATACAGCAGA 1553
QY 279 TTCAGAACAAACTTTTAAGTACAGCTAATGGAATGAAGATTAATGAAGACAACTTTAAA 338
Db 1554 CTTAGAAATGGTGGAAACACAACTTAATTTGCTGGAAGAAACACCAAGAAC 1613
QY 339 AAATTTAGATGACGAGAAATATGCTTATTAAGATTTAACTTATAATAATGCTAAAACTGT 398
Db 1614 ACCAGAAACCAAAAGAGAGAGTTACAACTCAAGTTAACTTAACTTTGCAGAA---TGG 1670
QY 399 TGAATTTAAAGTAGTAGCAGCTAGTGAAAAACAGTAGTTGTATCTTAGTGATGCGAAAAA 458
Db 1671 AAAACACAAACAGCAGAAATTCAAAGGAACATTTTGAAGAAACAAACAGCAGAAAGCTTACAG 1730

```

```

; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.406
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3279 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Peptostreptococcus asaccharolyticus
; STRAIN: 1018
; FEATURE:
; NAME/KEY: exon
; LOCATION: 103..3186
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 280
; OTHER INFORMATION: /product= "mature protein L"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 280..3183
; OTHER INFORMATION: /codon_start= 280
; OTHER INFORMATION: /product= "mature protein L"
; FEATURE:
; NAME/KEY: misc signal
; LOCATION: 208..279
; US-08-446-137B-1

```

QY 459 TAGTGCAAGATATAGCTCAAAAAATATGTTGTTGAAGACAAAGACTTAGAAAAATGCACCT 518
Db 1731 ATATGCAAGCTTATTAGCAAAAGTAATGTTGTAATACACAGCAGACTTAGAGATGGCGG 1790
QY 519 AAAAACTATATAATGCTCAGATTTCAGTAAACTGATAGTACTATCAAGTAGTCTTTA 578
Db 1791 ATACACTATCAACATCAAAATTTGCTGMAAAGAAACCAACAGGCGAAAAATCCAGGAATCAC 1850
QY 579 TCCAAAAGGAAGAGATTACAGGTTCTCAACTATAGAGCTACAAATATATATCAAGG 638
Db 1851 AATTGATGAATGGTTATTAAAGAAATGCTAAAGAAAGCAATCAAGAAATTTAAAGAACG 1910
QY 639 AACTGCAATGTTAAATACACAGCTAAATATTAACTCTAAATCTACTAGTAAGAGTAATTT 698
Db 1911 AGAATCACTTCGATTATTACTTCAGCTTAACTAATTAAGCAAAACAGCTTGAAGCGGT 1970
QY 699 AAGAGCTGCAAGTAGAAGAGTTTACAAAATTTGAATGCTAGTATTCTTAATACTACAATTT 758
Db 1971 AGAAGCAATTAAAGAACGAAATCTTAAAGCACACGCTGGAGAAAGAACACACAGAAATTA 2030
QY 759 AGCTGGTGATGACAGAAATACAAACAGCTATAGAGATAGTAAGTAAGTAATTT--ACATTA 815
Db 2031 AGATGATATGCAACATATGAAGAACGAGAGCAGCAGCTTAAAGAACTTTGAAATGA 2090
QY 816 TGAATGCGAGAAATCAGATCATTTCAGCTGATGTTTAAAGAGAAATGTTTAAATGTTGTATT 875
Db 2091 TGAATGTTAAACAGCATACGAAATAGTTTCAAGGTGCGAGCGGAAGATCTACTATGTAAT 2150
QY 876 AGTAGGTGCAAAATGCACTAGTAGATGATGTTGCGGCTCTTTAGCAGCAGAAAAAGA 935
Db 2151 AAGAGTTGAAGTTGCAGACGAGAAAGAACCCAGGTGAAGACACTCCAGAAAGTTCAAGAGG 2210
QY 936 TGCTCCACTATTATTAACTTCAAAAGATAAATAGATGCTGAGTAAATCTGAAATAAA 995
Db 2211 TTACGCAACTTACGAAGAACGAGAGCAGCAGCTTAAAGAGCAATTAAGAAAGATAAAGT 2270
QY 996 GAGAGTTTGTAGCTTAAAACTTCAACAGAAAGTAACAGGAAAAACAGTTTATATAGCTGG 1055
Db 2271 TAACATGTCATACGAGTAGTTTCAAGGTGCGAGCGGAAGATCTACTATGTTTAAAT 2330
QY 1056 TGGAGTTTAATAGTGTATCTAAGAAAGTTGTAACAGAAATTAAGATCAATGGGATTAAGT 1115
Db 2331 CGAAGATTAAGAAAGATGAACACAGCTGGAAGAACCCAGGGAACCCAGGAATCAAT 2390
QY 1116 TGAAGATCTCAGGTGATGATGATGAACTTTCTTTAAATATACAGGTGAAATAGG 1175
Db 2391 TGATGAATGGTTATTAAAGAAATGCTAAAGAGAGCGCAATCAAGAAATTAAGAAAGCAGG 2450
QY 1176 CTTAGATAATGATA 1189
Db 2451 AATCAGTTCTGACA 2464

RESULT 12
US-09-502-540-1357
; Sequence 1357, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1357
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Myxococcus xanthus

; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(612)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1357
Query Match 3.8%; Score 69.8; DB 4; Length 612;
Best Local Similarity 44.6%; Pred. No. 1.1e-05;
Matches 269; Conservative 0; Mismatches 334; Indels 0; Gaps 0;
QY 264 AGATAAAAAAGTTTATTTCAGACCAACTTTTAACTACAGCTTAATGGAAATGAAGATTATGT 323
Db 6 AAAAAAATTTAAAAAATTTTAACTACAGCTTAATGGAAATGAAGATTATGT 65
QY 324 AAGACCAACTTTAAAAAATTTTAACTACAGCTTAATGGAAATGAAGATTATGT 383
Db 66 AAAAAAATTTAAAAAATTTTAACTACAGCTTAATGGAAATGAAGATTATGT 125
QY 384 TAATGCTAAACCTGTTTGAATTTAAAGTAGTAGCAGCTAGTGAAAAACAGTAGTTGTATC 443
Db 126 AAAAAAATTTAAAAAATTTTAACTACAGCTTAATGGAAATGAAGATTATGT 185
QY 444 TAGTGATGCGAAAAATAGTGCAGAAAGATATAGCTGAAATATATGTTTGAACACCAAGA 503
Db 186 AAAAAAATTTAAAAAATTTTAACTACAGCTTAATGGAAATGAAGATTATGT 245
QY 504 CTTAGAAATGCACTAAAAAATTTTAACTACAGCTTAATGGAAATGAAGATTATGT 563
Db 246 ATAAAAAATTTAAAAAATTTTAACTACAGCTTAATGGAAATGAAGATTATGT 305
QY 564 TCAAGTAGTTCTTTTATCCAAAGAGAAAGAGATTACAAAGGTTTCTCAACTTATAGAGCTAC 623
Db 306 AAAAAAATTTAAAAAATTTTAACTACAGCTTAATGGAAATGAAGATTATGT 365
QY 624 AATTTAATGAAGAGGAGTGCATATGTAATACACAGTAATATTAATCTTAAATCTAC 683
Db 366 AAAAAAATTTAAAAAATTTTAACTACAGCTTAATGGAAATGAAGATTATGT 425
QY 684 TAGTAAGAGTAATTTTAAAGACTGCACTAGAGAGTTTACAAAAATTTGAATCTAGTTATTC 743
Db 426 AAAAAAATTTAAAAAATTTTAACTACAGCTTAATGGAAATGAAGATTATGT 485
QY 744 TAATACTTACAACTTTTAGCTGCTGATGACAGAAATACAAACAGCTATAGAGATAGTAAAGA 803
Db 486 AAAAAAATTTAAAAAATTTTAACTACAGCTTAATGGAAATGAAGATTATGT 545
QY 804 ATATTCAATATGATGGCGAGAAATCAGATCAATCAGCTGATGTTTAAAGAGATGTTAA 863
Db 546 AAAAAAATTTAAAAAATTTTAACTACAGCTTAATGGAAATGAAGATTATGT 605
QY 864 AAA 866
Db 606 AAA 608

RESULT 13
US-08-973-462-3
; Sequence 3, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-BRYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3

Db	987	CTGAAGTTAAAGCAGAAAGATTTCGAACCTGCTGTAGATGATAGTGTGGAACATTCATCAA	1046
Qy	707	CAGTAGAAGAGTTACAAAATTGGAATGCTAGTTATTCTTAATACTACAACTTTAGCTGGTG	766
Db	1047	GTGAAGTTGGAGAAAAGATCTCTGAACCTAGTAAAGAGAAAATACTCTCTGAAGTTAAAG	1106
Qy	767	ATCAGAGATACAAACAGCTATAGAGATAGTAAAGAAATATTCAATAATGATGGCGAGA	826
Db	1107	CGGAAGATTGCAACCTGCTGTAGATGGTGTAGAACATTCATCAAGTGAAGTTGGGA	1166
Qy	827	AATCAGATCATTCAGCTGATGTTAAAGAGAAATGTTAAATAATGTTATTAGTAGGTGCAA	886
Db	1167	AAAAAGTATCTGAACCTAGTAAAGAGGAAAGTACTCTCTGAAGTTAAAGCAGAAAGATTGC	1226
Qy	887	ATGCACTAGTAGATGGATTAGTTGGGCTCTTTAGCAGCAGCAAAAAGATGCTCCACTAT	946
Db	1227	AACCTGCTGTAGATG---ATAGTGTGGAACATTCATCAAGTGAAGTTGGGAGAAAAGTAT	1283
Qy	947	TATTAACTTCAAAGATAAATTAGATTGCTGTCAGTAAATCTGAAATAAAGAGAGTTTGTAG	1006
Db	1284	CTGAACCTAGTAAGAGGAAAATACTCTCTGAAGTTAGCAGCAAGATTGCAACCTGCTG	1343
Qy	1007	ACTTAAAACTTCAACAGAAAGTAAACAGGAAAAACAGTTTATATAGCTGGTGGAGTTAATA	1066
Db	1344	TAGATGCTAGTTAGAACATTCATCAAGTGAAGTTGGAGAAAAGTATCTGAAACTAGTA	1403
Qy	1067	GTGTATCTAAAGAGTTGTACAGAAATGAGATCAATGGGATTTAAAGTTGAAAGATTCT	1126
Db	1404	AAGAGGAAAGTACTCTCTGAAGTTAAAGCAGAAAGATTGCAACCTGCTGTAGATAGTA	1463
Qy	1127	CAGGTGATGATAGATAGTAAGTAATCTTTTAAATAAGCAGGTGAAATAGGCTTAGATAATG	1186
Db	1464	TAGAACATTCATCAAGTGAAGTTGGGAAAAGATATC---TGAAACTAGTAAAGAGGAAA	1520
Qy	1187	ATAAGGCTTATAGTTGGTGGAAACAGGATTAGCAGATGCCATGAGTATAGCTTCAGTTG	1246
Db	1521	GTACTCTGAAGTTAAAGCAGAGATTGTCACCTGCTGTAGATGGTGTAGAACATT	1580
Qy	1247	CTTCTACTAAATTAGATGTAATGGTGTGTAGATAGAACAAATGGACATGCTACTCCAA	1306
Db	1581	CATCAAGTGAAGTTGGAGAAAAGTATCTGAAACTAGTAAAGAGGAAAATACTCTCGAAG	1640
Qy	1307	TAGTTGTTAGATGGAAGCTGATAAATAATCTGATGACTTATAGATAGTT---TCTTAG	1363
Db	1641	TTAAAGCAGAAAGATTTCGAACCTGCTGTAGATGCTGTAGATGATGTAGAACATTCAAGTGAAG	1700
Qy	1364	GAAGCGCTGATAGATATAATAGTGGATTGCAAGTGTATCTGAAAAGATGGAAGAAG	1423
Db	1701	TTGAGAAAAGTATCTGAACTAGTAAAGAGGAAAATACTCTCTGAAGTTAAAGCGGAAG	1760
Qy	1424	CTATATCAGATGCTACTGGTAAAGCGGTTACAGAGTTAAAGGCGCAGTAGACAAGACA	1483
Db	1761	ATTTGCAACCTGCTGTAGATGGTGTAGTGTAGAACATTCATCAAGTGAAGTTGGAGAAAAG	1820
Qy	1484	CTAACTCTGAAGTTATAAACAATATTATGCTAATGATCTGAAATAGCTAAAGCTCGAG	1543
Db	1821	TATCTGAACTAGTAAAGAGAAAGTACTCTCTGAAGTTAAAGCGGAAAGATTGCAACCTG	1880
Qy	1544	TTTTAGATAAAGATTTCAGTGCTTCAAGTAGTGTAGCAGGAGTATTTAAATTTCTATGTAG	1603
Db	1881	CTGTAGATGGTGTGGACATTCATCAAGTGAAGTTGGGAGAAAAGTATCTGAGACTA	1940
Qy	1604	CTAAGATGGATCTACAAAAGAGATCAATTTAGTTGATGTCATTTAGCAGTAGGAGCTGTTG	1663
Db	1941	GTAAGAGAAAGTACTCTCTGAAGTTAAAGCGGAAGATTTGCAACCTGCTGTAGATGTA	2000
Qy	1664	CTG 1666	
Db	2001	GTG 2003	
RESULT 15			
US-08-323-170B-1			

Sequence 1, Application US/08323170B	
Patent No. 5733772	
GENERAL INFORMATION:	
APPLICANT: Williamson, Kim C.	
APPLICANT: Kaslow, David C.	
TITLE OF INVENTION: Cloning and Expression of Plasmodium	
TITLE OF INVENTION: faliciparum Transmission-Blocking Target Antigen, Pfs230	
NUMBER OF SEQUENCES: 4	
CORRESPONDENCE ADDRESS:	
ADDRESSEE: Townsend and Townsend and Crew LLP	
STREET: Two Embarcadero Center, 8th Floor	
CITY: San Francisco	
STATE: California	
COUNTRY: USA	
ZIP: 94111-3834	
COMPUTER READABLE FORM:	
MEDIUM TYPE: Floppy disk	
COMPUTER: IBM PC compatible	
OPERATING SYSTEM: PC-DOS/MS-DOS	
SOFTWARE: Patent In Release #1.0, Version #1.30	
CURRENT APPLICATION DATA:	
APPLICATION NUMBER: US/08/323,170B	
FILING DATE: 13-OCT-1994	
CLASSIFICATION: 424	
PRIOR APPLICATION DATA:	
APPLICATION NUMBER: US 08/010,409	
FILING DATE: 29-JAN-1993	
ATTORNEY/AGENT INFORMATION:	
NAME: Quine, Jonathan A.	
REGISTRATION NUMBER: P-41,261	
REFERENCE/DOCKET NUMBER: 015280-113100US	
TELECOMMUNICATION INFORMATION:	
TELEPHONE: (415) 576-0200	
TELEFAX: (415) 576-0300	
INFORMATION FOR SEQ ID NO: 1:	
SEQUENCE CHARACTERISTICS:	
LENGTH: 9636 base pairs	
TYPE: nucleic acid	
STRANDEDNESS: single	
TOPOLOGY: linear	
MOLECULE TYPE: DNA (genomic)	
FEATURE:	
NAME/KEY: CDS	
LOCATION: 149..9556	
US-08-323-170B-1	

Query Match	3.8%	Score 69.4;	DB 1;	Length 9636;
Best Local Similarity	42.5%	Pred. No. 2.7e-05;		
Matches	623;	Conservative	0;	Mismatches 821;
			Indels	21;
			Gaps	4;

Qy	370	GATTTAACTTATTAATAATGCTAAACCTGTTGAAATTAAGTAGTAGCAGCTAGTGAATAAA	429
Db	569	GTGTGTAGTAGTGATGCAATTCGAATGTACAAGAAAAAATATGATATTAATAAGAAAT	628
Qy	430	ACAGTAGTTCTACTAGTGATGCGAAAAATAGTGCAAAAAGATATAGCTGAAAAATATGTG	489
Db	629	ATAGATAAGAAAATAAATAATGATATACAAAATGTTGAGGAAAAAATACAAAGGATACA	688
Qy	490	TTTGAAGACAAAGACTTAGAAAAATGCACCTAAAACTATAAATGCCTCAGATTTTCAGTAAA	549
Db	689	TACGAAAAATAAGATTATGAAGTGATGATACACCTTATAGAAATGGTTTGATGATAATACA	748
Qy	550	ACTGATAGTTACTATCAAGTAGTCTTTATCCAAAGGAAAGAGATTTACAGGTTTCTCA	609
Db	749	AATGAAGAAAACTTTTACTAACTTTTAAAAAGGCTTGATGAAAAATATTTTCTTCA	808
Qy	610	ACTTATAGAGCTACAAATTAATGAAGGAATGTCATATGTTAATACACCAAGTAATATTA	669
Db	809	CCCAAGAAAAAAACTGTAGTACAAAAAACAATAAGTCTAATTTTTTATAAACAGT	868
Qy	670	ACTCTAAATCTACTAGTAAGAGTAATTTAAAGCTGCAGTAGAAGAGTTA-----CAA	723
Db	869	TCGTTGAAATATATATATATATATTTTAAACCCCTCGGATAGCTTTAACCTAGTAGTCGA	928

Qy	724	AAATTGAATCCTAGTTATTCTAACTACTACAACTTTTAGCTGGTGATCACAAGATACAAACA	783
Qy	1798	GTTTATAAACCAAAATGAAAGATTTAT	1822

Search completed: October 6, 2005, 10:25:02
Job time : 341 secs

D_b

674 GAGGAGGAAGATATGATTATGTCTTTTGTAGAAGTGGGGATCAAAACAGAACAA 1108

Db	1109	TTCAAGAGGAACATCAGGAAGAAGTAGGTCTGAATCTTCAGAGAAGGTTTTATGAT	1168
Ov	964	---AAATTAGATTCCGTCCAGTAAATCTCGAAATAAGAGAGAGTTTTAGACTTAAAAACTTCA	1020

1080
ACAGAACTAACAGGAAAAACAGTTTATATAGCTGGTGGAAGTTAATAGTAGTATCTAAAGAA

Qy

Qy 1081 GTTTGTAACAGCAATTAGACATCAATCGGATTAAGAAGTTGAAAG---ATTCTCAGGTGATCAT 1137

Db 1289 GTAGCTCAACAGCTAGCTGAAGAGAGGTAGGTGAAGAGGTAGGTGAAGAGAGGTAGGTCAAGAG 1348

[illegible]

Db	1409	GTAGGTGAACGGGTGAGGTGAAGAGGAGGTCAGAGAGGTCAGAGAA	1468
Ov	1258	TTTAGATCGTTAAATCGGTCTTCTAGATAGACAAATGCGCATGCTTACCTCCAAATAGCTCTTCTGTA	1317

DD 1469 GAGGGGAAATAGTGTAGATGAAAAAGAGAGGCGAGGATGAAATATATATCCATTTGGTGTATGAA 1328
 QY 1318 GATGGAAAGCTGATAAAAATATCTGATGACTTAGATAGTATTTCTTAGGAAGCGCTGATGTA 1377

Qy 1378 GATATAATAGTGGATTTCGAAGTGTATCTGAAAAGATCGAAGAAGCTATATCAGATGCT 1437

Db 1580 GATAAAAACAGATTTGTTTAAATTTATAGAAGGGGGTGAAGGGAGATGATATATAAAGTG 1639

QY 1438 ACTGGTAAAGCGGTTAAAGAGTAAAGCGGATAGACAAACATCTCTGAGGTT 1497

D5 1640 GATGGTTCCTCAAGTTTTATTAGATGATGATACAAATTAGTAGGTATCTAAAAACATACT 1699

Db	1700	GCACGAGATGGTGAATATGGTGAATATGGTGAACGTGTCGAAGATGCGAGAAATGTTTATA	1759
Dv	1558	TCAGGTGCTTTCAGTACTGATGATGACGAGAGATATTTAAATTTTCTATGCTAGCTAAAGATCGCATCT	1617

DB	QY
1760	ACAAAAGAAGTCAATTAAGTGTGATGCTTGGCGATATAAAGCTT 1677
1760	AAAAATAATTAGAAGTGTGTTACCAAGTGGTGGCAATACCAAGTGTAGTGTGTTGATGAGTTA 1819
1618	ACAAAAGAAGTCAATTAAGTGTGATGCTTGGCGATATAAAGCTT 1677

Qy 1678 GCTCCAGTTGATTAGCTACTGATCTTTTATCTTCTGATCAATCGGTTGCTATTAAGCAA 1737

Dh —————

1680 TCATATCATAAATATGTCATCTTAATATATACAAATAAAGAAATACGTTTGATTTTTACAGT 1939

QY
1738 GTTGATGAGAAATAATCTTAAAGATTTAAACAAGTTGGTCAGGAAATACCTAATCA 1797
Db
1940 CAATTAAACCAACAGAGAAGTGGTCCTTAAAGTAAAAAATGTGAAGTAAAGTTAATGAG 1999

	1798	1822
Qy	1798 GTTATAAACAAAATGAAAGATTTAT	1822
Db	2000 CCATTATAAAAGTAAAAATAATAT	2024

Search completed: October 6, 2005, 10:25:02
Job time : 341 secs

Result	Query No.	Score	Match	Query %	Length	DB	ID	Description	
C	1	113.8	6.2	1805	9	CL080711	CL080711	CH216-159	
	2	111.6	6.1	1784	9	CL081992	CL081992	CH216-165	
	3	106.2	5.8	2001	9	CL118787	CL118787	ISB1-72M2	
	4	104.6	5.7	1416	9	CL499036	CL499036	SAIL-662	
	5	104.2	5.7	1843	9	AG435185	AG435185	Mus muscu	
	6	104	5.7	1378	9	AG350209	AG350209	Mus muscu	
	7	103	5.6	1436	9	CL057761	CL057761	CH216-860	
C	8	102	5.6	1632	9	CL082569	CL082569	CH216-167	
	9	101.8	5.6	1376	9	CG747831	CG747831	PO41-3-B0	
C	10	101.8	5.6	1493	9	CL078589	CL078589	CH216-151	
	11	101.8	5.6	1506	9	AG278469	AG278469	Mus muscu	
C	12	101.8	5.6	1608	9	CL118721	CL118721	ISB1-72J8	
	13	101.6	5.6	1594	9	CL078613	CL078613	CH216-151	
	14	100.4	5.5	1162	9	CL077122	CL077122	CH216-143	
C	15	100.4	5.5	1519	9	AG386893	AG386893	Mus muscu	
	16	100.4	5.5	1531	9	CG748014	CG748014	PO41-4-B0	
C	17	100.4	5.5	1542	9	AG386981	AG386981	Mus muscu	
C	18	100.4	5.5	1566	9	CG757757	CG757757	P053-1-D0	
	19	100.2	5.5	1217	9	CL078190	CL078190	CH216-148	
C	20	100.2	5.5	1599	9	CL083840	CL083840	ISB1-2H14	
	21	99.8	5.5	1594	9	CL038406	CL038406	CH216-46A	
C	22	99.6	5.4	1842	9	CL044158	CL044158	CH216-59P	
C	23	99.4	5.4	1367	9	CL082650	CL082650	CH216-168	
	24	99.4	5.4	1434	9	AJ592058	AJ592058	Arabidops	

Db 1215 TAATAATAAAAAATATTGATTAAAAATAAAAAATATATATTAATAAATGAAAAATG 1156

Qy 146 ACTTAGTAAAGGATATTTTAGCGACGCAAAACTTAAACAAGCGTGCAAGTATTTTGTGAACA 205

Db 1155 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 1096

Qy 206 AGATACAAAGTTACTTCTATGATGCAAAAGAGAAAGATCTTCAACTCCACTGGAG 265

Db 1095 AATATATAATTATTAATAAATGAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 1036

Qy 266 ATAAAAAGTTTATTTCAGAACAACT - -TTAACTACAGCTAAATGGAAATGAAGATTATGT 323

Db 1035 AAAAAAATAAATATTTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 976

Qy 324 AAGACAACTTTAAAAAATTTAGATCGAGGAGATATGCTATTATATAGATTAACTTTATAA 383

Db 975 AAAAAAATAAAGAAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 917

Qy 384 TAATGCTAAACTGTTGAATTTAAAGTAGTAGCAGCTAGTGAAAAAACAAGTAGTTGTATC 443

Db 916 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 857

Qy 444 TAGTGATGCAAAATAGTGCAAAAGATATAGCTGAAAAATATGCTTTTGAAGACAAAGA 503

Db 856 AATAATGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 801

Qy 504 CTTAGAAATGCACTAAAACTATATAATGCCTCAGATTTTCAGTAAAACTGATAGTTACTA 563

Db 800 AAAAAAATAATAGATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 741

Qy 564 TCAGTAGTCTTTATTCGCAAGGAAAGAGATTAAGAGTTTCTCAACTTATAGAGTAC 623

Db 740 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 681

Qy 624 AATATTAAATGAAGAACTGCATATGTTAATACACAGTAATATTAACCTTAAATCTAC 683

Db 680 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 621

Qy 684 TAGTAAGAGTAATTTAAAGACTGCAGTAGAAGAGTTTACAAAAATTTGAATCTGATTTATTC 743

Db 620 TAAAGTGAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 561

Qy 744 TAATACTACAACCTTTAGCTGGTGATGACAGATACAAACAGCTATAGAGATAGTAAGA 803

Db 560 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 501

Qy 804 ATATTACAATAATGATGGCGAGAAATCAGATCAATCAGCTGATGTTTAAAGAGAAATGTTAA 863

Db 500 ATTATAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 441

Qy 864 AATGTTGTATTAGTAGGTGCAAAATGCATAGTAGATGATTAGTTGGCGCTCCTTTAGC 923

Db 440 TAATATTATATAAATAATGATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 381

Qy 924 AGCAGAAAAAGATGCTCCACTATTATTAACCTTCAAAAGATAAATTAGATTCTGCAGTAA 983

Db 380 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 321

Qy 984 ATCTGAAATAAAGAGAGTTTGTAGCTTAAAACTTCAACAGAAAGTAACAGGAAAAACAGT 1043

Db 320 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 261

Qy 1044 TTATATAGCTGGTGGAGTTTAATAGTGTATCTAAAGAAAGTTGTAACAGAAATTAGAATCAAT 1103

Db 260 TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 201

Qy 1104 GGGATTAAAGTTGAAGATTCTCAGGTGATGATGATGAACCTTCTTTTAAAAATAGC 1163

Db 200 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 141

Qy 1164 AGGTGAAATAGGCTTAGATAATGATAAGGCTTATGTAGTTGGTGGAAACAGGATTAGCAGA 1223

Db 140 AATTNANNANAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 81

Qy 1224 TGCCATAGTATAGCTTCAGTTGCTTCTACTAAATTAGATGTAATGTTGTTAGATAG 1283

Db 80 NNTAAAAAGATATTAAANNCCCTTTGAAAAAAGAAAAAGATAATTCGNATNAAANCT 21

Qy 1284 AACAATGGACA 1295

Db 20 AAAAAATTTAAAA 9

RESULT 2

CL081992

LOCUS

DEFINITION

CH216-165M9_RM4.1 CH216 Xenopus tropicalis genomic clone

CH216-165M9, genomic survey sequence.

CL081992

CL081992.1 GI:40537905

GSS.

Xenopus tropicalis (western clawed frog)

Xenopus tropicalis

Amphibia; Batrachia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 1784)

Krenitski,C., Carter,J., McPherson,J., Warren,W., Graves,T.,

Mardis,E. and Wilson,R.

A physical map of the xenopus tropicalis genome

Unpublished (2003)

Contact: Richard K Wilson

Genome Sequencing Center

Washington University School of Medicine

Email: submissions@watson.wustl.edu

Insert Length: 175000 Std Error: 0.00

Seq primer: RM4 ctcaagggcacgcgctcgagc

Class: BAC ends

High quality sequence start: 409

High quality sequence stop: 503.

Location/Qualifiers

1..1784

/organism="Xenopus tropicalis"

/mol_type="genomic DNA"

/strain="Nigerian frog"

/db_xref="taxon:8364"

/clone="CH216-165M9"

/sex="male"

/cell_line="Stock 248 F7A2, inbred N7"

/clone_lib="CH216"

/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis

BAC library"

ORIGIN

Query Match 6.1%; Score 111.6; DB 9; Length 1784;

Best Local Similarity 45.0%; Pred. No. 2.9e-12;

Matches 497; Conservative 0; Mismatches 604; Indels 3; Gaps 2;

Qy 87 AATATCACTACAGATGGTACAAATGATAAGTATACAGTATCAATACAAATACAAAGCTAGTGA 146

Db 480 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 539

Qy 147 CTTAGTAAGGATATTTTAGCAGCACAAACTTACACAGCTGCGAGTTATTTTGAACAA 206

Db 540 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 599

Qy 207 AGATACAAAAAGTTACTTTCTATGATGCAAAATGAGAAAGATTTCTTCAACTCGGAGA 266

Db 600 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 659

Qy 267 TAAAAAGTTTATTCAGAACAACTTTTAACTACAGCTAATGGAAATGAAGATTATGTAAA 326

Db 660 AATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 719

Qy 327 GCACTTTTAAAAATTTAGATGCGAGGAGATATGCTATTATAGATTTAACCTTATAATAA 386

Db 720 AA 779

Qy 387 TGTCTAACTGTTGAAATTAAGTACTAGCAGCTAGTGAAACACAGTAGTTGTACTAG 446

Db 780 AA 839

Qy 447 TGATGCGAAAAATAGTGCAGAAAGATATAGCTGAAAAAT-ATGTGTTTGAAGCAAAAGCT 505

Db 840 GAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 899

Qy 506 TAGAAAAAGCACTAAAGAACTAATAATGCGCTCAGATTTTCAGTAAAGTGTAGTACTATC 565

Db 900 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 959

Qy 566 AAGTAGTCTTTATCCAAAGGAAAGAGATTAAGAGGTTTCTCACTTATAGAGCTACAA 625

Db 960 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1019

Qy 626 ATTATTAATGAAGGAAGCTGCATATGTTGTAATACACAGTAAATTAATCTTAAATCTACTA 685

Db 1020 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1079

Qy 686 GTAAGAGTAAATTAAGAGCTGAGTAGAGAGTTACAAAAATTTGAATGCTAGTTATCTA 745

Db 1080 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1139

Qy 746 ATACTCAACTTTAGCTGGTGTGACAGATACAAACAGCTATAGAGATAGTAAAGAT 805

Db 1140 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1199

Qy 806 ATTACAATATGATGCGAGAAATCAGATCAITTCAGCTGATGTTAAAGAGAAATGTTAAAA 865

Db 1200 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1259

Qy 866 ATGTTGTTATGAGTGGAAGTGCAGTACTAGTAGAGTATGATGATGATGCGGCTCTTTAGCAG 925

Db 1260 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1319

Qy 926 CAGAAAAAGATGCTCACTATTATTAATTAACCTCAAGAGTAAATTTAGATTCGTCAGTAAAT 985

Db 1320 AATAAATAAAGAGATATAGAAAGAAAAAATAATATAAATAGAAAAAATAAATAAATAA 1379

Qy 986 CTGAATAAAGAGAGTTTATAGCTTAAAGAACTTCAACAGAGTAAACAGGAAAAACAGTTT 1045

Db 1380 AGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1439

Qy 1046 ATATAGCTGGTGAGTTAATAGTGTATCTAAAGAGTTGTAACAGAAATTAAGATCAATGG 1105

Db 1440 AATAAATAAATGA--AAAGAAAAAATAAATAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1497

Qy 1106 GATTAAAGTTGAAGATTTCTCAGTGATGATAGATATGAAGCTTCTTTAAAGATAGCAG 1165

Db 1498 TAAAAAAGAAAAAAGAGAAAAAATAAATAAAGAAAAAATAAAGATGTATATAAATAAAGAA 1557

Qy 1166 GTCAATATAGCTTAGATAATGATA 1189

Db 1558 AAAAAATATGAAGAAAGATAATA 1581

RESULT 3

LOCUS CL118787

DEFINITION ISB1-72M23_T7.2 ISB1 Xenopus tropicalis genomic clone ISB1-72M23, genomic survey sequence.

ACCESSION CL118787

VERSION CL118787.1

KEYWORDS GI:40612422

SOURCE GSS.

ORGANISM Xenopus tropicalis (western clawed frog)

ORGANISM Xenopus tropicalis

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae; Xenopodinae; Xenopus; Silurana.

REFERENCE 1 (bases 1 to 2001)

AUTHORS TITLE JOURNAL COMMENT

Krenitski,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
A physical map of the xenopus tropicalis genome
Unpublished (2003)
Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: T7 TAATACGACTCACTATAGG
Class: BAC ends
High quality sequence start: 384
High quality sequence stop: 783.
Location/Qualifiers
1. .2001
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone_lib="ISB1-72M23"
/note="Vector: pBelOBAC11; ISB-1 Xenopus tropicalis BAC Library Segment 1"

FEATURES source

ORIGIN

Query Match 5.8%; Score 106.2; DB 9; Length 2001;
Best Local Similarity 43.4%; Pred. No. 3.7e-11;
Matches 471; Conservative 0; Mismatches 612; Indels 2; Gaps 1;

Qy 87 AATATCACTACAAGATGGTACAAATGATAGTATACAGTATCAATACTTAAAGCTAGTGA 146

Db 99 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 158

Qy 147 CTTAGTAAAGGATATTTTAGCAGCAGCAAACTTAAACACAGGTGCAGTTATTTTGAACAA 206

Db 159 CAATAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 218

Qy 207 AGATACAAAAGTTACTTTCTATGATGCAATGAGAAAGATTTCTTCAACTCCAACTGAGA 266

Db 219 CTTTATAAAGCGGNNNNNNNNNNAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 278

Qy 267 TAAAAAGTTTTCAGAACAACTTTAACTCAGCTAATGGAATGGAATGATATGTAAA 326

Db 279 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 338

Qy 327 GACAACCTTTAAAGAAATTTAGATGCAGGAGATATGCTATTATAGATTTAACTTATATA 386

Db 339 AGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 398

Qy 387 TGTAAACCTGTTGAAATTTAAAGTAGTAGCAGCTAGTGAACAAAGTGTGTCTCTAG 446

Db 399 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 458

Qy 447 TGTGCGAAAAATAGTGCAGAAAGATATAGCTGAAAAATATGTGTTTGAAGACAAAGCTT 506

Db 459 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 518

Qy 507 AGAAAAATGCATAAAACCTATAATGCTCAGATTTTCAAGTAACTGATAGTTTACTATCA 566

Db 519 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 578

Qy 567 AGTAGTTCTTTTCCAAAAAGGAAAGATTTACAGGTTTCTCAAGTGTCTTCACTTATAGAGCTACAA 626

Db 579 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 638

Qy 627 TTATATGAAGGAAGCTGCATATGTTATATACACAGTAAATTAATTAATTAATTAATTA 686

Db 639 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 698

Qy 687 TAAGAGTAAATTTAAAGCTGCTAGTAGAGGTTTACAAAAATTTGAATGCTAGTTTCTTAA 746

Db 699 AGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 758

Qy 747 TACTACAACCTTTAGCTGGTGATGACAGATACAAACAGCTTATAGAGATTAAGTAAAGATA 806


```

sequence.
ACCESSION AG350209
VERSION AG350209.1 GI:47923519
KEYWORDS GSS.
SOURCE Mus musculus molossinus
ORGANISM Mus musculus molossinus

REFERENCE
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE BAC end Sequences of Library MSMg01
JOURNAL Unpublished
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY : pBACe3.6
Vector : EcoRI
R.Site 1 : EcoRI
R.Site 2 : EcoRI
Location/Qualifiers
1. .1378
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/cl_xref="taxon:57486"
/clone="MSMg01-146M02.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN
Query Match 5.7%; Score 104; DB 9; Length 1378;
Best Local Similarity 45.9%; Pred. No. 9.9e-11;
Matches 430; Conservative 0; Mismatches 501; Indels 6; Gaps 2;

QY 108 AATGATAGTACGTATCAATCTAAAGCTAGTCTAGTAAAGGATATTTAGC 167
Db 1346 ATAATAAATTAATAATATATCTATATAAAAAATATAATAATTAATAATTAATACT 1287
QY 168 AGCAGAAAACCTTAACAACAGGTGCAGTTATTTTGACAAAGATACAAAA--GTTACTTT 224
Db 1286 ATAAATAATATATTAATTAATTAATAACAATAATTAATAACCAATACATAAATAA 1227
QY 225 CTATGATGCAATGAGAAAGATTCTTCAACTCCAAGTGGAGATAAAAAAGTTTATTCAGA 284
Db 1226 ATATATATTAAATATATAAAAAAATAATATAAAATAATATAATAATAATTAATAAAA 1167
QY 285 ACAAACTTTAATACAGCTAATGGAATGAAGA---TTATGTGAAGACAACTTTAAAAAA 341
Db 1166 ATAAAAATTTAAATATATAAAAAATATATAAAAAAATATAAAAAAATAAAAAATAATAT 1107
QY 342 TTTAGATGCGAGGAATATGCTATATAGATTAACTTATAATATGCTAAAACTGTTGA 401
Db 1106 ATAAATAAAATATAATACATATAATAATAATAATAATAATAATAATAATAATAATA 1047
QY 402 AATTAAGTAGTACGACGTAGTGAAAAAACAAGTAGTTGATCTAGTGATGCGAAAAATAG 461
Db 1046 AAAAAATTTAAAAATTTAAATTTAAATAATTTAAATAATATAAAAAAATAAAAAATAAAAA 987

```

```

QY 462 TGCRAAAGATATAGCTGAAAAATATGTCTTTGAAGACACAAAGACTTTAGAAAAATGCACATAA 521
Db 986 TAATAAAAAAATAATAAAAAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 927
QY 522 AACTATAAATGCCTCAGATTTTCAGTAAAACTGATAGTTACTATCAAGTAGTTCCTTTATCC 581
Db 926 AAATAACACATAATTTAAATATAATAAAAAAATAAATAAATAAATAAATAAATAAATAAATA 867
QY 582 AAAAGGAAGAGATTACAAAGTTTCTCAACTTATAGAGCTACAAATTTATTAATGAAGGAC 641
Db 866 TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAT 807
QY 642 TGCATATGGTATACACCAGTAATAATTAACTCTAAAACTACTACTAGAGAGTAATTTAAA 701
Db 806 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 747
QY 702 GACTGCAGTAGAAGATTTCACAAAAATTCGAATGCTAGTTATTCTTAATACTACACAACTTTAGC 761
Db 746 TTATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 687
QY 762 TGGTGATGACAGATACAAAAGCTATAGAGATAAGTAAAGAAATTTACAATAATGATGG 821
Db 686 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 627
QY 822 CGAGAAATCAGATCAATTCAGCTGATGTTAAAGAGAGATGTTAAAAATGTTGTTAGTAGG 881
Db 626 AAAAAAATTTATAATAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 567
QY 882 TGCAAAATGCACCTAGTAGATGATTTGCGGCTCCTTTAGCAGCAGAAAAAGATGCTCC 941
Db 566 ATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 507
QY 942 ACTATTATTAATCTCAAAAGATAAATAGATTGCTGAGTAACTGAGTAAATCTGAAATAAGAGAGT 1001
Db 506 AATAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 447
QY 1002 TTTAGACTTAAAAAATCTCAACAGAGATTAACAGGAAAA 1038
Db 446 GTAATTATTATAAATATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 410

RESULT 7
LOCUS CL057761
DEFINITION CH216-8609 RMI.1 CH216 xenopus tropicalis genomic clone CH216-8609,
genomic survey sequence.
ACCESSION CL057761
VERSION CL057761.1 GI:40513674
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1436)
AUTHORS Krenitski,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: RM1 TACGACTCACTATAGGGAGA
Class: BAC ends
High quality sequence start: 220
High quality sequence stop: 1095.
Location/Qualifiers
1. .1436
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"

FEATURES
source

```

```
/db_xref="taxon:8364"
/clone="CH216-8609"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/notes="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
BAC library"

ORIGIN

Query Match      5.6%; Score 103; DB 9; Length 1436;
Best Local Similarity 43.4%; Pred. No. 1.6e-10;
Matches 472; Conservative 0; Mismatches 615; Indels 0; Gaps 0;

QY 87 AATATCACTACAGAGTGGTACAAATGATAGTATACAGTATCAAACTTAACACAGAGTGAAGTAACTAGTGA 146
DB 1229 AAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1288

QY 1047 TATAGTCTGGTGGAGTTAATAGTATCTAAAGAACTTGTAAACAGAACTTAGAATCAATGCG 1106
DB 1289 GAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAA 1348

QY 1107 ATTAAAAAGTTGAAGATTTCTCAGGTGATCATAGATATGAAACTTCTTTAAAAATAGCAGG 1166
DB 1349 AAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 1408

QY 1167 TGAATA 1173
DB 1409 AGAAAAA 1415

RESULT 8
LOCUS      CL082569
DEFINITION CH216-167P5 Sp6.1 CH216 Xenopus tropicalis genomic clone
ACCESSION  CL082569
VERSION     CL082569.1 GI:40538482
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
ORGANISM    Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1 (bases 1 to 1632)
AUTHORS     Krenitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
TITLE        A physical map of the xenopus tropicalis genome
JOURNAL      Unpublished (2003)
COMMENT      Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000 Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 1062
            High quality sequence stop: 1114.

FEATURES             Location/Qualifiers
     source            1..1632
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-167P5"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /notes="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"

ORIGIN

Query Match      5.6%; Score 102; DB 9; Length 1632;
Best Local Similarity 42.7%; Pred. No. 2.6e-10;
Matches 471; Conservative 0; Mismatches 633; Indels 0; Gaps 0;

QY 87 AATATCACTACAGAGTGGTACAAATGATAGTATACAGTATCAAACTTAACACAGAGTGAAGTAACTAGTGA 146
DB 430 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 489

QY 147 CTTAGTAAAGGATATTTTAGCAGCACAAAACCTTAAACAGAGTGCAGTTATTTTGAACAA 206
DB 490 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 549

QY 207 AGATACAAAAGTTACTTTCTATGATGCAATGAGAAAGATCTTCACTCCAACTGGAGA 266
DB 550 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 609
```


QY 627 TTATAAGTAAAGCAATGCGATATGCTAATACACAGTAATATTAACTCTAAATCTACTAG 686
 Db 632 AA 573
 QY 687 TAAGAGTAATTTAAAGCTGCGATAGAGAGTTTACAAAATTGGAATGCTAGTTTCTTAA 746
 Db 572 AA 513
 QY 747 TACTACAACTTTAGCTGGTGCATGACAGATACAAACAGCTATAGAGTAAGTAAAGCAATA 806
 Db 512 ANAA 453
 QY 807 TTCAATAATATGATGGCGAGAAATCAGATCATCTTCAGCTGTATGTTTAAAGAGAAATTTAAAAA 866
 Db 452 AA 393
 QY 867 TGTGTATTAGTAGGTGCAATGCATGATAGATGGAATTAGTTGGCGCTCTTTAGCAGC 926
 Db 392 AA 333
 QY 927 AGAAAAAGATGCTCCACTATTATTAACTTCAAAAGATTAATTTAGATTGCTCAGTAAATC 986
 Db 332 AA 273
 QY 987 TGAATAAAGCAGAGTTTTCAGCTTAAAACTTCAACAGAGTAAACAGGAAAAACA 1041
 Db 272 AA 218

RESULT 10
 CL078589
 LOCUS
 DEFINITION
 CH216-151K5_Sp5.1 CH216 Xenopus tropicalis genomic clone
 CH216-151K5, genomic survey sequence.
 ACCESSION
 VERSION
 CL078589.1 GI:40534502
 KEYWORDS
 SOURCE
 ORGANISM
 Xenopus tropicalis (western clawed frog)
 Xenopus tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Xenopus; Silurana.
 1 (bases 1 to 1493)
 Krenitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
 Mardis, E. and Wilson, R.
 A physical map of the xenopus tropicalis genome
 Unpublished (2003)
 Contact: Richard K Wilson
 Genome Sequencing Center
 Washington University School of Medicine
 Email: submissions@wustl.edu
 Insert Length: 175000 Std Error: 0.00
 Seq primer: Sp5 atctgcgcttcgatact
 Class: BAC ends
 High quality sequence start: 1009
 High quality sequence stop: 1070.
 Location/Qualifiers
 1..1493
 /organism="Xenopus tropicalis"
 /mol_type="genomic DNA"
 /strain="Nigerian frog"
 /db_xref="taxon:8364"
 /clone="CH216-151K5"
 /sex="male"
 /cell_line="Stock 248 F7A2, inbred N7"
 /clone_lib="CH216"
 /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
 BAC library"

FEATURES

source
 1..1493
 /organism="Xenopus tropicalis"
 /mol_type="genomic DNA"
 /strain="Nigerian frog"
 /db_xref="taxon:8364"
 /clone="CH216-151K5"
 /sex="male"
 /cell_line="Stock 248 F7A2, inbred N7"
 /clone_lib="CH216"
 /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
 BAC library"

Query Match 5.6%; Score 101.8; DB 9; Length 1493;
 Best Local Similarity 42.4%; Pred. No. 2.8e-10;
 Matches 464; Conservative 0; Mismatches 631; Indels 0; Gaps 0;

ORIGIN

QY 1116 TGAAGATTCTCAGGTGATGATGATGAACTTCTTTAAATACGAGGTGAATAGG 1175
 Db 1393 AAATAAATAATAAANAATATAAANAATAAATAAATAAATAAATAAATAAANA 1452

QY 96 ACAGATGCTACAAATGATTAAGTATACAGTATCAATACTAAAGCTAGTACCTTAGTAAA 155
 Db 373 AA 432
 QY 156 GGATATTTTTCAGCAGCAGCAAACTTTAAACACAGCTGCGATGTTTTCAGCAAGATACAAA 215
 Db 433 AA 492
 QY 216 AGTTTCTTCTATGATGCAAAATGAGAAAGATTCTTCAACTCCAACTGGAGATAAAAAAGT 275
 Db 493 AA 552
 QY 276 TTATTCAGAACAACTTTTAACTACAGCTAATCGAAATGAAGATTATGTTAAGCAACTTTT 335
 Db 553 AA 612
 QY 336 AAAAAATTTAGATGCGGAGATATGCTATTATAGTTTAACTTATATAAATGCTTAAAC 395
 Db 613 NAAA 672
 QY 396 TGTTCGAAATTAAGTAGTAGCAGCTAGTCGAAAAACAGTAGTTGTATCTAGTATGCGAA 455
 Db 673 AA 732
 QY 456 AATAGTGCAGAAAGATATAGCTGAAAAATATGTGTTTGAAGCAAAAGACTTAGAAATGC 515
 Db 733 AA 792
 QY 516 ACTAAAACTATAAATGCTCAGATTTTCAGTAAACTGATGTTACTTCTCAAGTATGTTCT 575
 Db 793 AA 852
 QY 576 TTATCCAAAAGGAAGAGATTTCAAGGTTTCTCACTTATAGAGCTACAAATTTATATGA 635
 Db 853 AA 912
 QY 636 AGGAATCGCATATGGTAATACACAGTAATATTAACTCTAAATCTACTAGTAGAGTAA 695
 Db 913 AA 972
 QY 696 TTTTAAAGCTGCGATGAGAGAGTACAAAATTTGAATGCTAGTTTCTTAATCTACAC 755
 Db 973 TAAA 1032
 QY 756 TTTTACCTGCTGATGACAGATACAAACAGCTATAGAGATAAGTAAAGAAATTTACATAA 815
 Db 1033 AA 1092
 QY 816 TGATGCGAGAAATCAGATCATTCAGCTGATGTTTAAAGAGAAATGTTTAAATGTTGATT 875
 Db 1093 AA 1152
 QY 876 AGTAGTGCAATGCTAGTAGATGGAATTAGTTGGCGCTCTTTAGCAGCAGAAAAGA 935
 Db 1153 ATAAAAAAAAAAAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1212
 QY 936 TGCTCCACTATTATTAACTTCAAAAGTAAATTTAGATTTCGTCACTAAATCTGAAATAAA 995
 Db 1213 AAAAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1272
 QY 996 GAGAGTTTTAGACTTAAAAACTTCAACAGAGTAAACAGGAAACAGTTTATATAGCTGG 1055
 Db 1273 AAAAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1332
 QY 1056 TGGAGTTTAAATGATGTTCTTAAAGAGTTGTAACAGAAATTAAGATCAATGGGATTAAGT 1115
 Db 1333 AAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1392
 QY 1116 TGAAGATTCTCAGGTGATGATGATGAACTTCTTTAAATACGAGGTGAATAGG 1175
 Db 1393 AAATAAATAATAAANAATATAAANAATAAATAAATAAATAAATAAATAAANA 1452

```

QY 1176 CTTAGATATGATAA 1190
DB 1453 NTAATAATAATAAA 1467

RESULT 11
AG278469/c
LOCUS Mus musculus molossinus DNA, clone:MSMg01-048M18.T7, Genomic survey
DEFINITION sequence.
ACCESSION AG278469
VERSION AG278469.1 GI:47851346
KEYWORDS GSS.
SOURCE Mus musculus molossinus
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
BAC end Sequences of Library MSMg01
Unpublished
2 (bases 1 to 1506)
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
Direct Submission
Submitted (17-NOV-2003) Mashira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsuri-mi-ku, Yokohama, Kanagawa 230-0045, Japan
[E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170]
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : T7
LIBRARY : pBAC3.6
Vector : EcoRI
R.Site 1 : EcoRI
R.Site 2 : EcoRI
FEATURES
source Location/Qualifiers
1..1506
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-048M18.T7"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN
Query Match 5.6%; Score 101.8; DB 9; Length 1506;
Best Local Similarity 44.7%; Pred. No. 2.8e-10;
Matches 489; Conservative 0; Mismatches 597; Indels 9; Gaps 2;

QY 80 CAGATGTAATCACTACAGAGTGGTACAAATGATAGTATACAGTACCAATCTAAAG 139
DB 1244 CAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1185

QY 140 CTAGTGACTTAGTAAAGGATATTTTAGCAGCACAATACTTAAACAAGGTGCAGTTATTT 199
DB 1184 AAAAAAATTAAAGAAATAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1125

QY 200 TGAACAAGATACAAAGTTACTTTCTATGATGCCAATGAGAAAGATTTCTTCACTCCAA 259
DB 1124 ATAAAAATATATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1065

QY 260 CTGCAGATAAAAAGTTTATTTCAGAACAACTTTAACTACAGCTAATGGAATGAAGATT 319
DB 1064 ATAAATATAAATTAAATATAAATAAATAAATAAATAAATAAATAAATAAATAA 1005

```

```

QY 320 ATGTAAAGACAACTTTTAAAAAATTTAGATGCGAGGAGATATGCTATTATTAGATTAACTT 379
DB 1004 AAATAAAA-----TATATAAATAAATAAATAAATAAATAAATAAATAAATAA 951

QY 380 ATAATAAGTCTAAAACTGTTGAAATTTAAAGTAGTAGCAGCTAGTGCAGAAACAGTAGTTG 439
DB 950 TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 891

QY 440 TATCTAGTAGTCGAAAAATAGTCGAAAGATATATAGCTGAAATAATATGTGTTTGAAGACA 499
DB 890 AAAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 831

QY 500 AAGACTTAGAAAAATGCCTAAATACTATAAATGCCTCAGATTTCTAGTAAAACTGATAGTT 559
DB 830 ATCAAGGGTAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 771

QY 560 ACTATCAAGTACTGTTCTTTATCCAAAGGAAAGAGATATCAAGGTTTCTCAACTTATAGAG 619
DB 770 AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 711

QY 620 CTACAAATTAATAATCAAGGAACTGCATATGTAATACACAGTAAATATTAATCTTAAAT 679
DB 710 ATAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 654

QY 680 CTACTAGTAAGAGTAAATTTAAAGACTGCAGTAGAAGAGTTACAAAAATTTGAATGCTAGTT 739
DB 653 ATAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 594

QY 740 ATTCTTAATCTACTCAACTTTTAGCTGCTGATGACAGATACAAACAGCTATAGAGATAAGTA 799
DB 593 AAAAAAAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 534

QY 800 AAGAATATTTCAATAATGATGCGGAGAAATCAGATCATTTTCAGCTGATGTTTAAAGGAATG 859
DB 533 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 474

QY 860 TTAATAATGTTGTTATTAGTAGTGCAATGCACTAGTAGTAGTATGCTGCGGCTCCTT 919
DB 473 AAAAAATAAAGGGGAGAAAAATATAAATAAATAAATAAATAAATAAATAA 414

QY 920 TAGCAGCAGAAAAAGATGCTCCACTATTATTAACTTTCAAAAGATAAATTTAGATTTCGTCAG 979
DB 413 AAAAAATAATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 354

QY 980 TAAAAATCTGAATAAAGAGAGTTTTTAGACTTTAAAAAACTTTCAACAGAGTAAACAGGAAAA 1039
DB 353 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 294

QY 1040 CAGTTTTATATAGCTGCTGGAGTTAATAGTGTATCTTAAGAAGTTGTAACAGAAATTAGAT 1099
DB 293 AACAATACAAAAAACAACAAAAATAAATAAATAAATAAATAAATAAATAA 234

QY 1100 CAATGGGATTAAGTTTGAAGATTTCTCAGGTGATGATAGATATGAACCTCTTTAAAAA 1159
DB 233 ACTTTACCTTTTATCTACTTAAACAATCATATATTTGTTGAAATTTGGAACGGAATAACAATT 174

QY 1160 TAGCAGGTGAAATAG 1174
DB 173 TCACACAGGAACAG 159

RESULT 12
CL118721
LOCUS ISBI-72J8_T7.1 ISBI Xenopus tropicalis genomic clone ISBI-72J8,
DEFINITION Genomic survey sequence.
ACCESSION CL118721
VERSION CL118721.1 GI:40612356
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```



```
Best Local Similarity 42.5%; Pred. No. 3.1e-10;
Matches 470; Conservative 0; Mismatches 634; Indels 2; Gaps 1;

QY 87 AATATCACTACAGATGGTACAAATGATAGTATACAGTATCAAAATACATAAGCTAGTGA 146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 350 AAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 409

QY 147 CTTAGTAAAGGATATTTTAGCAGCACAAACTTAAACAAGGTGCGATTATTTTGACAA 206
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 410 NNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 469

QY 207 AGATACAAAGGTACTCTTCTATGATGCAATGAGAAAGATTCTTCAACTCCACTGGAGA 266
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 470 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 529

QY 267 TAAAAAGTTTATTTCAGAACAACTTTAACTACAGCTAATGAAATGAAGATTATGTAA 326
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 530 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 589

QY 327 GACAACTTTAAANATTTAGTCAGGAGAATATGCTATTATAGATTTAACTTTAATAA 386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 590 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 649

QY 387 TGTCTAAACTGTTGAAATTAAGTACTAGCAGCTAGTGAANNAACAGTAGTTGTCTAG 446
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 650 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 709

QY 447 TGATGCGAAANATAGTGCAAAGATATAGCTGAAANATATGTGTTTGAAGACAAAGACTT 506
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 710 AANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 769

QY 507 AGAAATGCACTTAAANCTATTAATGCTCAGATTTTCAGTAAACAGTAGTACTATCA 566
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 770 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 829

QY 567 AGTAGTCTTTATCCAAAGGAAGAGATTACAAAGTTTCTCAACTTATAGAGCTACAA 626
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 830 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 889

QY 627 TTATAATGAAGGAATGCGATATGTTAATACACAGTAATATTTAACTC--TAAANATCTACT 684
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 890 AAAAAANNAANNAATTAANNAANNAANNAANNAANNAATTAANNAATTAANNAANNA 949

QY 685 AGTAAGAGTAAATTTAAAGACTGCGATGAGAGAGTTACAAANATTTGAATGCTAGTTATCT 744
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 950 AAAAAANNAANATTAANNAANNAANNAANNAANNAANNAANNAANNAANNA 1009

QY 745 AATACTACAACTTTAGCTGCTGATGACAGAAATACAAACAGCTATACAGATAAGTAAAGAA 804
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1010 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 1069

QY 805 TATTACAATAATGATGCGAGAAATCAGATCAATTCAGCTGATGTTTAAAGAGAATGTTAA 864
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1070 AATAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 1129

QY 865 AATGTTGTATTAGTAGTGAATGCACTAGTAGATGGAATTTAGTTCGGGCTCCTTTAGCA 924
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1130 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 1189

QY 925 GCAGAAAGAGATGCTCCACTATTATTAACTTCAAAAGATAAATTTAGATTTCGTCAGTAA 984
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1190 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 1249

QY 985 TCTGAAATGAAGAGAGTTTTAGACTTTAAANCTTCAACAGAGTAACAGGAAANNAACAGTT 1044
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1250 AATAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 1309

QY 1045 TATATAGCTGGTGGAGTTAATAGTGTATCTAAAGAGAGTTGTAAACAGAAATTAGAATCAATG 1104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1310 AAAAAANNAATTAANNAANNAANNAANNAANNAANNAANNAANNAANNA 1369

QY 1105 GGATTAAGAGTTCTCAAGGTGATGATAGATATGAAATCTTCTTTAAANATAGCA 1164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 1370 AAAAAAAAAATAAANNAANNAANNAATTAANNAANNAATTAANNAANNAATTAANNA 1429

QY 1165 GGTGAATAGGCTTAGATTAATGATAA 1190
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1430 AATTAANNAANNAANNAATTAANNAATTA 1455

RESULT 14
CL077122 1162 bp DNA linear GSS 31-DEC-2003
LOCUS CH216-143B5 Sp6.1 CH216 Xenopus tropicalis genomic clone
DEFINITION CH216-143B5_ genomic survey sequence.
ACCESSION CL077122
VERSION CL077122.1 GI:40533035
SOURCE GSS.
ORGANISM Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1162)
AUTHORS Krenitski,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 757
High quality sequence stop: 808.
Location/Qualifiers
1..1162
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strains="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-143B5"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTABAC2.1; CHORI-216 Xenopus tropicalis
BAC library"

ORIGIN
Query Match 5.5%; Score 100.4; DB 9; Length 1162;
Best Local Similarity 42.7%; Pred. No. 5.3e-10;
Matches 471; Conservative 0; Mismatches 630; Indels 3; Gaps 1;

QY 87 AATATCACTACAAAGTGGTACAAATGATAAGTATACAGTATCAAAATACATAAGCTAGTGA 146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38 ANTCTCCAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 97

QY 147 CTTAGTAAAGGATATTTTAGCAGCACAAACTTAAACAAGGTGCGATTATTTTGACAA 206
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 98 AAAAAANNAATCTTAANNAANNAANNAANNAANNAANNAANNAANNAANNA 157

QY 207 AGATACAAAGTTACTTTCTATGATGCAATGAGAAAGATTCTTCAACTCCACTGGAGA 266
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 158 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 217

QY 267 TAAAAAGTTTATTTCAGAACAACTTTTAACCTACAGCTATGGAATGAAGATTATGTAA 326
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 218 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 277

QY 327 GACAACTTTAAANATTTTAGTCAGGAGAATATGCTATTATAGATTTAACTTTAATAA 386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 278 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 337

QY 387 TCGTAAACTGTTGAATTTAAAGTAGTAGCAGCTAGTGAAANNAACAGTAGTTGTATCTAG 446
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM nucleic - nucleic search, using sw model

Run on: October 6, 2005, 06:06:36 ; Search time 3690 Seconds
(without alignments)
3454.108 Million cell updates/sec

Title: US-10-068-870A-4

Perfect score: 1830

Sequence: 1 atgaaaaaagaatttagc.....tgaagattattagatg 1830

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 8443130 seqs, 3482420727 residues

Total number of hits satisfying chosen parameters: 16886260

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09D_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US10J_PUBCOMB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US11B_PUBCOMB.seq:*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1830	100.0	1830	14	US-10-068-870-2
2	542	29.6	2271	14	US-10-068-870-4
3	535.6	29.3	2271	18	US-10-239-610-6
4	503.4	27.5	2217	14	US-10-068-870-6
5	483.2	26.4	2145	18	US-10-239-610-4
6	480	26.2	2145	14	US-10-068-870-7
7	471.8	25.8	2157	14	US-10-068-870-1
					Sequence 2, Appli
					Sequence 4, Appli
					Sequence 6, Appli
					Sequence 6, Appli
					Sequence 4, Appli
					Sequence 7, Appli
					Sequence 1, Appli

8	471.8	25.8	2158	14	US-10-068-870-3	Sequence 3, Appli
9	471.8	25.8	2158	14	US-10-068-870-8	Sequence 8, Appli
10	471.8	25.8	2160	18	US-10-239-610-5	Sequence 5, Appli
11	470.2	25.7	2158	14	US-10-068-870-5	Sequence 5, Appli
12	145.2	7.9	1893	17	US-10-282-122A-16859	Sequence 16859, A
13	104	5.7	1503	17	US-10-282-122A-16782	Sequence 16782, A
14	99.6	5.4	1830	17	US-10-282-122A-16845	Sequence 16845, A
15	95.6	5.2	5361	9	US-09-742-096-2	Sequence 2, Appli
16	95.6	5.2	5361	9	US-09-742-096-1	Sequence 1, Appli
17	94.8	5.2	5528	18	US-10-415-253-1	Sequence 1, Appli
18	92	5.0	4997	17	US-10-282-122A-35506	Sequence 35506, A
19	88.4	4.8	5253	17	US-10-282-122A-17449	Sequence 17449, A
20	87.6	4.8	7442	18	US-10-221-714A-409	Sequence 409, App
21	86.2	4.7	7446	17	US-10-282-122A-7578	Sequence 7578, Ap
22	85.2	4.7	7434	9	US-09-815-242-4761	Sequence 4761, Ap
23	85.2	4.7	7437	9	US-10-470-048B-201	Sequence 201, App
24	85.2	4.7	8155	8	US-09-815-243-8869	Sequence 8869, Ap
25	85.2	4.7	8155	8	US-08-781-986A-63	Sequence 63, Appl
26	85.2	4.7	8155	18	US-10-329-624-63	Sequence 63, Appl
27	84.8	4.6	1300	19	US-10-668-749A-1	Sequence 1, Appli
28	84.8	4.6	4911	21	US-10-470-048B-427	Sequence 427, App
29	82.2	4.5	2370	11	US-09-754-947-2	Sequence 2, Appli
30	81.8	4.5	6292	18	US-10-221-714A-461	Sequence 461, App
31	81.4	4.4	1243	20	US-10-425-115-172717	Sequence 172717,
32	80.6	4.4	2409	17	US-10-282-122A-17248	Sequence 17248, A
33	79	4.3	1204	19	US-10-437-963-77858	Sequence 77858, A
34	76.4	4.2	921	20	US-10-425-115-38710	Sequence 38710, A
35	74.8	4.1	778	20	US-10-363-345A-2179	Sequence 2179, Ap
36	74.8	4.1	778	20	US-10-363-345A-2180	Sequence 2180, Ap
37	74.8	4.1	778	21	US-10-363-483A-2179	Sequence 2179, Ap
38	74.8	4.1	778	21	US-10-363-483A-2180	Sequence 2180, Ap
39	74.8	4.1	1039	20	US-10-425-115-58802	Sequence 58802, A
40	74.2	4.1	5000	21	US-10-706-635-25	Sequence 25, Appl
41	74	4.0	1062	20	US-10-425-115-120013	Sequence 120013,
42	73.8	4.0	15548	15	US-10-311-455-2128	Sequence 2128, Ap
43	73.8	4.0	3673778	16	US-10-312-841-1	Sequence 1, Appli
44	73.6	4.0	1845	17	US-10-282-122A-16828	Sequence 16828, A
45	73.2	4.0	1830	21	US-10-706-635-67	Sequence 67, Appl

ALIGNMENTS:

RESULT 1

US-10-068-870-2
; Sequence 2, Application US/10068870

; Publication No. US20030054009A1

; GENERAL INFORMATION:

; APPLICANT: The Provost, Fellows & Scholars of the College of the Holy and Undivided

; APPLICANT: Trinity of Queen Elizabeth, near Dublin

; TITLE OF INVENTION: C.difficile vaccine

; FILE REFERENCE: TRI002/C/WO

; CURRENT APPLICATION NUMBER: US/10/068,870

; CURRENT FILING DATE: 2002-02-11

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2

; LENGTH: 1830

; TYPE: DNA

; ORGANISM: Clostridium difficile

; US-10-068-870-2

Query Match 100.0%; Score 1830; DB 14; Length 1830;

Best Local Similarity 100.0%; Pred. No. 1.3e-277;

Matches 1830; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAAAAAAGAAATTTAGCAATGGCTATGCGAGCTGTTACTGTAGTAGGTTCTGCTCT 60

Db 1 ATGAAAAAAGAAATTTAGCAATGGCTATGCGAGCTGTTACTGTAGTAGGTTCTGCTCT 60

Qy 61 CCAGTTTTTCAGCAGCTTCAGATGTTATATCACTACAGATGCTACAAATGATAAGTAT 120

Db 61 CCAGTTTTTCAGCAGCTTCAGATGTTATATCACTACAGATGCTACAAATGATAAGTAT 120

```
QY 121 ACAGTATCAAACTAAAGCTAGTGAAGTAAAGGATATTTTACGACGACAAAACCTTA 180
Db 121 ACAGTATCAAACTAAAGCTAGTGAAGTAAAGGATATTTTACGACGACAAAACCTTA 180
QY 181 ACAACAGGTGCAGTATTTTGAACAAAGATACAAAAGTTACTTTCTATGATGCAAAATGAG 240
Db 181 ACAACAGGTGCAGTATTTTGAACAAAGATACAAAAGTTACTTTCTATGATGCAAAATGAG 240
QY 241 AAGATTTCTTCAACTCAACTCGAGATAAAAAAGTTTATTCAGAACAAAACCTTTAACTACA 300
Db 241 AAGATTTCTTCAACTCAACTCGAGATAAAAAAGTTTATTCAGAACAAAACCTTTAACTACA 300
QY 301 GCTAATGGAAATCAAGATTTATGTAAGACAACTTTTAAAGATTTAGATCGAGAGATAT 360
Db 301 GCTAATGGAAATCAAGATTTATGTAAGACAACTTTTAAAGATTTATGATCGAGAGATAT 360
QY 361 GCTATTATAGATTTAACTTTATATAATGCTTAAACCTGTTGAAATTTAAAGTAGTAGCAGCT 420
Db 361 GCTATTATAGATTTAACTTTATATAATGCTTAAACCTGTTGAAATTTAAAGTAGTAGCAGCT 420
QY 421 AGTGAACAAAACAGTAGTTGTATCTAGTGATGCGAAAAATAGTGCAAAAGATATAGCTGAA 480
Db 421 AGTGAACAAAACAGTAGTTGTATCTAGTGATGCGAAAAATAGTGCAAAAGATATAGCTGAA 480
QY 481 AATATGTGTTGAACACAAAGACTTAGAAATGCACTTAAACACTTAAATGCTTCAGAT 540
Db 481 AATATGTGTTGAACACAAAGACTTAGAAATGCACTTAAACACTTAAATGCTTCAGAT 540
QY 541 TTCAGTAAAACTGATAGTTACTATCAAGTAGTTCTTTTATCCAAAGGAAAGAGATATACAA 600
Db 541 TTCAGTAAAACTGATAGTTACTATCAAGTAGTTCTTTTATCCAAAGGAAAGAGATATACAA 600
QY 601 GGTTCCTCAACTTATAGAGCTACAAATATATGAAGGAAGTGCATATGGTAATACACCA 660
Db 601 GGTTCCTCAACTTATAGAGCTACAAATATATGAAGGAAGTGCATATGGTAATACACCA 660
QY 661 GTATATTAACCTTAAATCTACTAGTAGAGTAAATTTTAAAGCTGCAGTAGAAGGTTA 720
Db 661 GTATATTAACCTTAAATCTACTAGTAGAGTAAATTTTAAAGCTGCAGTAGAAGGTTA 720
QY 721 CAAAAATGAATCTAGTTATTTCTAATACTACAACTTTTAGCTGGTGATGACAGAAATACAA 780
Db 721 CAAAAATGAATCTAGTTATTTCTAATACTACAACTTTTAGCTGGTGATGACAGAAATACAA 780
QY 781 ACAGCTATAGAGATAAGTAAAGAAATATCAATATGATGGCGAGAAATCAGATCAATCA 840
Db 781 ACAGCTATAGAGATAAGTAAAGAAATATCAATATGATGGCGAGAAATCAGATCAATCA 840
QY 841 GCTGATGTTAAAGAGAAATGTTAAAGATGTTGATTTAGTAGGTGCAAAATGCAGTAGAT 900
Db 841 GCTGATGTTAAAGAGAAATGTTAAAGATGTTGATTTAGTAGGTGCAAAATGCAGTAGAT 900
QY 901 GGATTAGTTGCGGCTCCTTTAGCAGCAGAAAAAGATGCTCCACTATTATTAACTTCAAAA 960
Db 901 GGATTAGTTGCGGCTCCTTTAGCAGCAGAAAAAGATGCTCCACTATTATTAACTTCAAAA 960
QY 961 GATAAATTAGATTCGTGAGTAAATCTGAAATTAAGAGAGTTTTAGACTTAAAGAACTTCA 1020
Db 961 GATAAATTAGATTCGTGAGTAAATCTGAAATTAAGAGAGTTTTAGACTTAAAGAACTTCA 1020
QY 1021 ACAGAGTAAACGAGAAAAACAGTTTATATAGCTGGTGGAGTTAAATAGTGTATCTAAGAA 1080
Db 1021 ACAGAGTAAACGAGAAAAACAGTTTATATAGCTGGTGGAGTTAAATAGTGTATCTAAGAA 1080
QY 1081 GTTGTAACAGAAATTTAGAACTCAATGGGATTTAAAGTTTGAAGATTTCTCAGGTGATGATA 1140
Db 1081 GTTGTAACAGAAATTTAGAACTCAATGGGATTTAAAGTTTGAAGATTTCTCAGGTGATGATA 1140
QY 1141 TATGAAACTCTTTTAAAAATAGCAGGTGAAATAGGCTTATAGATAATGATAAGGCTTATGTA 1200
Db 1141 TATGAAACTCTTTTAAAAATAGCAGGTGAAATAGGCTTATAGATAATGATAAGGCTTATGTA 1200
```

```
QY 1201 GTTGTGGAAACAGGATTAGCAGATGCCATGAGTATAGCTTCCAGTTGCTTCTACTAAATTA 1260
Db 1201 GTTGTGGAAACAGGATTAGCAGATGCCATGAGTATAGCTTCCAGTTGCTTCTACTAAATTA 1260
QY 1261 GATGTTAATGGTGTGTTAGATAGAAACAAATGACATGCTTACTCCAAATAGTTGTTGTAGAT 1320
Db 1261 GATGTTAATGGTGTGTTAGATAGAAACAAATGACATGCTTACTCCAAATAGTTGTTGTAGAT 1320
QY 1321 GGAAAGCTGATAAAATATCTGATGACTTATAGTAGTTTCTTAGGAAGCGCTGATGTAGAT 1380
Db 1321 GGAAAGCTGATAAAATATCTGATGACTTATAGTAGTTTCTTAGGAAGCGCTGATGTAGAT 1380
QY 1381 ATAAATAGTGGATTTGCAAGTCTATCTGAAAGATGGAAGAGCTATATCAGATGCTACT 1440
Db 1381 ATAAATAGTGGATTTGCAAGTCTATCTGAAAGATGGAAGAGCTATATCAGATGCTACT 1440
QY 1441 GGTAAAGGCGTTTACAAGAGTTTAAAGCGCACATAGACAAACACTCTCAAGTTTATA 1500
Db 1441 GGTAAAGGCGTTTACAAGAGTTTAAAGCGCACATAGACAAACACTCTCAAGTTTATA 1500
QY 1501 AAAACATATTTATGCTTAATGATACTGAAATAGCTTAAAGCTGCAGTTTATAGATAAAGATTC 1560
Db 1501 AAAACATATTTATGCTTAATGATACTGAAATAGCTTAAAGCTGCAGTTTATAGATAAAGATTC 1560
QY 1561 GGTGCTTCAAGTAGTGCAGGAGTATTTAATTTCTATGTAGCTTAAAGATGGATCTACA 1620
Db 1561 GGTGCTTCAAGTAGTGCAGGAGTATTTAATTTCTATGTAGCTTAAAGATGGATCTACA 1620
QY 1621 AAAGAAGATCAATTTAGTTGATGCATTTAGCAGTAGGAGCTGTTTCTGGATATAAACTTGCT 1680
Db 1621 AAAGAAGATCAATTTAGTTGATGCATTTAGCAGTAGGAGCTGTTTCTGGATATAAACTTGCT 1680
QY 1681 CAGTTGTTATGACTGATGATTTCTTTATCTTCTGATCAATCGTTGCTTATAGCAAAAGTT 1740
Db 1681 CAGTTGTTATGACTGATGATTTCTTTATCTTCTGATCAATCGTTGCTTATAGCAAAAGTT 1740
QY 1741 GTAGAGAAAAATATTTCTAAAGATTTAAACAAAGTTGGTCAAGGAATAGCTAATTTCAAGTT 1800
Db 1741 GTAGAGAAAAATATTTCTAAAGATTTAAACAAAGTTGGTCAAGGAATAGCTAATTTCAAGTT 1800
QY 1801 ATAAACAAAATGAAAGATTTTATTAGATATG 1830
Db 1801 ATAAACAAAATGAAAGATTTTATTAGATATG 1830
```

RESULT 2

US-10-068-870-4

; Sequence 4, Application US/10068870

; Publication No. US20030054009A1

; GENERAL INFORMATION:

; APPLICANT: The Provost, Fellows & Scholars of the College of the Holy and Undivided

; APPLICANT: Trinity of Queen Elizabeth, near Dublin

; TITLE OF INVENTION: C.difficile vaccine

; FILE REFERENCE: TRI002/C/WO

; CURRENT APPLICATION NUMBER: US/10/068,870

; CURRENT FILING DATE: 2002-02-11

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 4

; LENGTH: 2271

; TYPE: DNA

; ORGANISM: Clostridium difficile

US-10-068-870-4

Query Match 29.6%; Score 542; DB 14; Length 2271;

Best Local Similarity 59.4%; Pred. No. 1.6e-75;

Matches 1117; Conservative 0; Mismatches 670; Indels 93; Gaps 8;

QY 32 CAGCTGTTACTGCTAGTAGGTTCTGCTGCTCCAGTTTTCAGCAGCAGCTTCAGATGTAATAT 91

Db 401 CAGCAGATGCTATAAATGCTGGAAATCTTCAGCAGATGGTGTGTTTACAAATACTGGAG 460

QY 92 CACTACAAGATGGTCAAAATGATATACAGTATCAATACTAAAGCTAGTACTTAG 151

461	CTGCTAGTGGTCTCTAGTGACAAAATTCAGCAGGAACAAAACTTGCAATGTCAGCTATTT	520
152	TAAAGATATTTTAGCAGCACAAAACCTTAACAACAGGTGCAGTTATTTTGAACAAGATA	211
521	TTGACACAGCATATACAGATTTCATCTGAACTCGCGTTAAGATTACTATAAAGCAGATA	580
212	CAAAAGTTACT---TTCTATGATGCAAAATGAGAAAGATTCTTCAACTCCAACCTGGAGATA	268
581	TGATCATACTAAATTTGGTAAAGCAGGTGAGACAACCTATTCACTGGGCTTACATTTG	640
269	AAAAAGTTTATTCAGAACAAAACCTTTAACTACAGCTAATGGAATGAAGATTATGTTAAAGA	328
641	AAGATGGGTCTACAGAAAAAAATTTGTTAAATAGGGGACAGTGTATATATAGATATATACTA	700
329	CAACTTTAAAA---AATTTAGATGCGAGGAGNAATGCTATTATAGATTTTAACTTATAATA	385
701	AAGCTCTTAAACTTACTGTGTTTCTGGAAGTAAAGCACTGTTAAGTTTGGCTGAAAAAA	760
386	ATGCTAAAAACGTGTG-----AAATTTAAAGTAGTAGCAGCTAGTGCAAA	427
761	CACCAAGTGCAGTGTTCACACAGTAATAACAAGCTTAGAATAATAAATGCTAAAGAAG	820
428	AAAACAGTGTGTATCTTAGTGTGCGAAAAATATAGTGCAAAAGATATAGCTGAAAAATATG	487
821	AAACAATAGATATTGACGCTAGTCTTAGTAAAAACAGCAACAAGATTAGCTTAAAAAATATG	880
488	TGTTTGNAGACAAGACTTTAGAAAAATGCCTAAAACTATAAATGCTCAGATTTTCAGTA	547
881	TATTTTAATAAACTGATTTTAAATCTCTTTATAAAGTATTAATAATGGAGATGAAGCAGATA	940
548	AAACTGATAGTTACTATCAAGTAGTTCTTTTATCCAAAAGCAAGAGATTACAAGGTTTCT	607
941	CTAATGGATTAAATAGAGAAGTTAGTGGAAAAATATCAAGTAGTTCTTTATCCAGAAGAA	1000
608	CAACTTATAGAGCTACAAAATTTAT---AATGAAGGAACCTGCATATGTTAATACACCAGTAA	664
1001	AAAGAGTTACAACTAAGAGTGTGCAAAAGCGTTCAATTCGTGATGAAAAATCCACCAAGTTA	1060
665	TATTAACCTTAAAACTACTAGTAAAGAGTAATTTTAAAGACTGCAGTAGAAGAGTTACAAA	724
1061	AATTTAACTCTTAAGTCAAGTAAAGAAGAAAGCACTTTAAAGAGATTATGTGGATGATTAAAGAA	1120
725	AATTGAATGCTAGTTATTTCTTAATACTACAACTTTAGCTGTGTGATGACAGAGAATACAAACAG	784
1121	CATATAAATATGGATATTTCAATGCTATAGAGTAGCAGAGAGAGATAGATAGAACTG	1180
785	CTATAGAGATAAGTAAAGAAATTTTACAAATATATGATGGCGAGAAATCAGATCAATTCAGCTG	844
1181	CAATAGCATTAAGTCAAAAATATTTAATCTGTATGATGAAAA-----TGCTA	1228
845	ATGTTAAAGAGATGTTTAAAAATGTTGTTATTTAGTAGGTGCNAATGCCTAGTAGATCGAT	904
1229	TATTTAGAGATTTCAGTTTGATAATGTAGTTATTTGGTTGGAGGAAATGCAATAGTTGATGGAC	1288
905	TAGTTGGGCTCCCTTTTAGCAGCAGAAAAAGATGCTCCAATATTATTAATTAACCTTCAAAAGATA	964
1289	TTGTAGCTTCTCTTTAGCTTCTGAAAAGAAAGCTTCTTTATTTATTTAATTTCAAAAGATA	1348
965	AATTAGATTTCGTCACTAGTAAAAATCTGAAAATAAAGAGAGTTTTAGACTTTAAAAAATTTCAACAG	1024
1349	AATTTAGATTCAAGCGTAAAGCTGAAATAAAGAGAGTTATGAAATAAAGAGTAGTCAACACAG	1408
1025	AAGTA---ACAGGAAAAACAGTTTATATAGCTGCTGGGTTTAAATAGTGTATCTTAAAGAG	1081
1409	GTATAAAATACTTCAAAGAAAGCTTTTATTTAGCTGTGGAGTTTAAATCTTATATCTTAAAGAG	1468
1082	TTGTAAACAGAAATTAAGATCAATGGATTAAAAGTTTGAAGATTCTCAGGTGATCATAGAT	1141
1469	TAGAAAATGAAATTTAAAGATATGGACTTAAAGTTTAAAGATTAGCAGGAGATGATAGAT	1528
1142	ATGAACTTCTTTTAAAAATAGCAGGTGAAATAGGCTTAGATTAATGATGAAGCTTTATGTAG	1201

RESIT.T 3

RESULTS
IIS-10-239-610-6

US-10-239-610-8
: Semence 6. Application US/10239610

: Publication No. US20040039165A1

; PUBLICATION NO.: US20
: GENERAL INFORMATION:
: GENERAL INFORMATION:

; GENERAL INFORMATION: ; APPLICANT: Imperial College of Science Technology and Medicine

APPLICANT: Imperial College of Science Technology and Medicine
TITLE OF INVENTION: CLOSTRIDIUM DIFFICILE POLYPEPTIDES AND USES THEREOF

; TITLE OF INVENTION: CLOSURE
: FILE REFERENCE: Y012 2 0008

FILE REFERENCE: Y00Z 2 00089
CURRENT APPLICATION NUMBER: IIS/10/239.610

; CURRENT APPLICATION NUMBER: US/10/239,61
 ; CURRENT FILING DATE: 2002-08-24

; CURRENT FILING DATE: 2002-09-24
; PCT APPLICATION NUMBER: PCT/GB01/01305

; PRIOR APPLICATION NUMBER: PCT/JP2001/03-32
 PRIOR FILING DATE: 2001-03-32

; PRIOR FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: CB 0007253 7

; PRIOR APPLICATION NUMBER: GB 2000003234

; PRIOR FILING DATE: 2000-

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE:

; SEQ ID NO 6

; LENGTH: 2

TYPE: DNA

ORGANISM: C

Query Match		29.3%;	Score 535.6;	DB 18;	Length 2271;
Best Local Similarity		59.2%;	Pred. No. 1.6e-74;		
Matches 1113;		Conservative	0;	Mismatches 674;	Indels 93;
				Gaps	8;
QY	32	CAGCTGTTACTGTAGTGGTCTGTCTCCAGTCTTTTTCAGCAGCTTCAGATGTAATAT	91		
DB	401	CAGCAGATGTAATTTGCTGGAACTCTTCAGCAGATGGTGTGTACAAATCTGGAG	460		
QY	92	CACTACAAGATGGTACAATAGATAACAGTATCAAGTATCAAACTAAGCTAGTGAATTAG	151		
DB	461	CTGCTAGTGGTCTTACTAGACAAATTCAGCAGAACTTCAGCAGGACAACTTGCATGTGAGCTATT	520		
QY	152	TAAAGGATTTTAGCAGCACAACTTAACACAGGTGCAGTATTTTGAACAAAGATA	211		
DB	521	TTGACACAGCATATACAGATTCATCTGAACTCGCGTTAAGATTTACTATANAAGCAGATA	580		
QY	212	CAAAAGTTACT---TTCTATGATGCAAAATGAGAAAGATTCTTCAACTCCAACTGGAGATA	268		
DB	581	TGAATGATATAAATTTGGTAAAGCAGGTGAGCACTTATTCACCTGGGCTTACATTTG	640		
QY	269	AAAAAGTTTATTCAGAACTTTAACTACAGCTAATGGAATGAAGATTATGTAAGA	328		
DB	641	AAAGTGGTCTACAGAAAAAATTTGTTAAATTTAGGGACAGTGATATTAAGATATAACTA	700		
QY	329	CAACTTTAAA---AATTTAGATGACAGGAATATGCTATTATAGATTTAACTTAATA	385		
DB	701	AAGCTCTTAACCTTACTGTGTCTTGGAAAGTAAGCAACTGTTAAGTTTGTGAAAAAA	760		
QY	386	ATGCTAAAACTGTG-----AAATTAAGTAGTAGCAGCTAGTGAAA	427		
DB	761	CACCAAGTCCAGTGTTCACACAGTAATAACAAGCTTAGAATAATAATGCTAAAGAG	820		
QY	428	AAACAGTAGTGTATCTAGTAGTGCAGAAAAATAGTGCAGAAAGATATAGCTGAAAAATAG	487		
DB	821	AAACATATAGATATTGACGCTAGTCTTAGTAAAAACAGCAGCAAGATTAGCTAAAAAATAG	880		
QY	488	TGTTTGAACACAAAGACTTAGAAAAATGCACATAAACTATAATGCCTCAGATTTTCAGTA	547		
DB	881	TATTTAATAAACTGAATTTAAATACCTTTTATAAAGTATTAAATGGAGATGAAGCAGATA	940		
QY	548	AAACTGATAGTTACTATCAAGTAGTTCTTTTATCCAAAAGGAAAGAGATTACAAGGTTTCT	607		
DB	941	CTAATGGATTATAGAGAAGTTAGTGGAAATATCAAGTAGTCTTTTATCCAGAAGNA	1000		
QY	608	CAACTTATAGAGCTACAAATTTAT---AATGAAGGAACCTGCATATGGTAAATACACAGTAA	664		
DB	1001	AAAGAGTTACAACTAAGAGTGTGCAAGGGCTTCAATTCCTGATGAAATTCACCAAGTTA	1060		
QY	665	TATTAACTTAAATCTACTAGTAAAGTAATTTAAAGACTGCAGTAGAAGAGTTACAAA	724		
DB	1061	AAATTAACCTTAAGTCAGATAAGAAGAAAGACTTAAAGAGTATTGTGGATGATTTAAGAA	1120		
QY	725	AATTTGAATCTAGTTATTTCTAATCTACAACTTTTAGCTGGTATGACAGAAATCAAAAACAG	784		
DB	1121	CATATAAATAGTATTTCAATGCTATAGAGTAGCAGGAGAGATAGATAGAACTG	1180		
QY	785	CTATAGAGATAAGTAAAGAAATTTACAATAATGATGGCGAGAAATCAGATCAATCAGCTG	844		
DB	1181	CAATAGCATTAAGTCAAAAATATTATAACTCTGATGATGAAAA-----TGCTA	1228		
QY	845	ATGTTAAAGAGAAATGTTAAAAATGTTTATTAGTAGGTGCAAAATGCACGTAGATGGAT	904		
DB	1229	TATTTAGAGATTCAGTTGATAATGATGATGTTGGTGGAGAAATGCAATAGTTGATGAC	1288		
QY	905	TAGTTGCGGCTCTTTTAGCAGCAGAAAAAGATGCTCCACTATTATTAACTTCAAAAGATA	964		
DB	1289	TTGTAGCTTCTCTTTAGCTCTGGAAGAAAGCTCCITTTATTATTAACTTCAAAAGATA	1348		
QY	965	AATTAGATTGCTGCTAAATCTGAAATTAAGAGAGTTTGTAGCTTAAACTTCAACAG	1024		
DB	1349	AATTAGATTCAAGCGTAAAGCTGAAATTAAGAGAGATTATGAATATAAAGAGTACAACAG	1408		
QY	1025	AAAGTA---ACAGAAAAACAGTTTATATAGTGGTGGAGTTAATAGTGTATCAAGAGAG	1081		

DB	1409	GTATAAATACCTTCAAGAAAGTTTTTATTAGCTGGTGGAGTTAATTTCTATATCTAAAGAG	1468		
QY	1082	TTGTAAACAGAAATTAGAATCAATGGGATTAAAGATTGAAGATTCTCAGGTGATGATAGAT	1141		
DB	1469	TAGAAAATGAATTTAAAGATATGGGACTTAAAGTTTCAAGATTAGCAGGAGATGATAGAT	1528		
QY	1142	ATGAAACTTCTTTTAAAAATAGCAGGTGAAATAGGGCTTAGAATAATGATAAAGGCTTAGTAG	1201		
DB	1529	ATGAAACTTCTTAAAAATAGCTGATGAAGTAGGTCTTGATAAATGATAAAGCATTTGTAG	1588		
QY	1202	TTGTGGGAACAGGATTAGCAGATGCCATGAGTATAGCTTTCAGTCTTCTACTAATAATTAG	1261		
DB	1589	TTGGAGGAACAGGATTAGCAGATGCCATGAGTATAGCTTCCAGTTCGATCTCAATTAAGAA	1648		
QY	1262	ATGCTAAATGGTGTCTTAGATAGAACAAATGGACATGCTACTCCAATAGTTGTTGTAGATG	1321		
DB	1649	ATGCTAAATGGTAAATGGATTTAGCTGATGGTGTGCTACACCAATAGTTAGTTGTAGATG	1708		
QY	1322	GAAGAGCTGATAAATAATCTGATGATCTTAGATAGTTTCTTTAGGAAGCGCTGATGATAGTA	1381		
DB	1709	GAAGAGCTTAAACTATAAATGATGATGTAAAGATTCTTCTAGATGATTCAACAAGTTGATA	1768		
QY	1382	TAATAGGTGGATTTTCCAAAGTGTATCTGAAAGATGGAAGAGCTATATCAGATGCTACTG	1441		
DB	1769	TAATAGGTGGAGAAAACAGTGTATCTAAAGATGTTGAAATGCAATAGATGCTACAG	1828		
QY	1442	GTAAAGGGGTTTACAGAGTTAAAGCGCAGATAGACAAGACACTAACTCTGAAAGTTATAA	1501		
DB	1829	GTAAATCTCCAGATAGATATAGTGGAGATGATAGACAGCAACTAATGCAAAAGTTATAA	1888		
QY	1502	AAACATATTTAGCTAATGATCTGAAATAGCTTAAAGCTGCAGTTTTTAGATAAAGATTGAG	1561		
DB	1889	AGAATCTCTTATTATCAAGATAACTTAAATAATGATAAAAAAGTAGTTAATTTCTTTG	1948		
QY	1562	GTGCTTCAAGTAGTGCAGAGTATTTAATTTCTATGTAGCTAAAGATGGATCTACAA	1621		
DB	1949	TAGCTAAAGATGGTCTTACTAAAGAGATCAATAGTTGATGCTTTAGCAGCAGCTCCAG	2008		
QY	1622	A-----AGAGATCAATTTAGTTG	1639		
DB	2009	TTGCAGCAAACTTTTGGTGAACCTCTTAATCTGATGGTAAGCCAGTAGTAAAGATGGTA	2068		
QY	1640	ATGCATTTAGCAGTAGGAGCTGTGCTGGATATAAA-----CTTGCTCCAGTTGTAT	1690		
DB	2069	AGTATTAACTGGTCTGATAATGATAAAAAATAAATTAGTATCTCCAGCACCCTATAGTAT	2128		
QY	1691	TAGCTACTGATTTCTTTCTCTGATCAATCGTGTGCTATAGCAAAAGTTGTAGGAGAAA	1750		
DB	2129	TAGCTACTGATTTCTTTCTCTGATCAATCGTGTGCTATAGCAAAAGTTGTAGGAGAAA	2188		
QY	1751	AATATTTCAAGATTTTAAACACAAAGTTGCTCAAGGAATAGCTAATTTCAAGTTATAAACA	1810		
DB	2189	ATAATGGAGAAAACCTTAGTTCAGTTGGTAAAGGTATAGCTACTTCACTTATAAACAAT	2248		
QY	1811	TGAAGATTTTATAGATATG	1830		
DB	2249	TAAAGATTTTATAGTATG	2268		

RESULT 4

US-10-068-870-6
; Sequence 6, Application US/10068870
; Publication No. US20030054009A1
; GENERAL INFORMATION:
; APPLICANT: The Provost, Fellows & Scholars of the College of the Holy and Undivided
; TRINITY OF QUEEN ELIZABETH, near Dublin
; TITLE OF INVENTION: C.difficile vaccine
; FILE REFERENCE: TRI002/C/WO
; CURRENT APPLICATION NUMBER: US/10/068,870
; CURRENT FILING DATE: 2002-02-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1


```

; SEQ ID NO 6
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Clostridium difficile
US-10-068-870-6

Query Match      27.5%; Score 503.4; DB 14; Length 2217;
Best Local Similarity 59.6%; Pred. No. 1.8e-69; Indels 27; Gaps 5;
Matches 950; Conservative 0; Mismatches 616;

QY 250 TCAACTCCAACTGGAGATAAAGTTTATTCAGAACAACTTTAACTACAGCTAATGGA 309
DB 640 TCTGTTCTTACTACAGCTTAACTAACTGCTGATACAACTGCAACAAACAGATGTAAT 699
QY 310 AATGAAGATTATGTAAGACAACTTTAAAGAAATTTAGATGCAGGAGAAATATGCTATTATA 369
DB 700 ATTCTGATGTTATGATGCTATTAATTTAATGTTACTGATACAGATTAGTGANTCCCA 759
QY 370 GATTTAACTTATAATAATGCTAAACTGTTGAAATTTAAAGTAGTAGCAGCTAGTGA AAA 429
DB 760 GCTGGTTCACTGCTTCTACTCTTAGAGCAAGTATTAAGTANTAAATGCMAAAGAA 819
QY 430 ACAGTAGTTGTATCTAGTGCAGGAAAGAAATAGTGCMAAAGATATAGCTGAAAGATATGTG 489
DB 820 TCTATAGATGTTGATTTCAAGTTCACATAGAACAGCTGAGATTTAGCTGAAAGATATGTA 879
QY 490 TTTGAAGACAAAGACTTTAGAAATGCACTAAAGAACTATTAATGCTCCTCAGATTTC----- 543
DB 880 TTTAAACCAAGAGATGTGAATAAATCTTATGAGGCACTGACTGATTTATATAAAGAGGT 939
QY 544 AGTAAAGCTATAGTATCTATCAAGTAGTTCTTTTATCCAAAGGAAAGAGATTAACAAGGT 603
DB 940 ATAACAAGTAACTCTTATCACTCAAGATGTTGGAATAATATCAAGTGTGTTTATTTGCTCAA 999
QY 604 TTCTCACTTTATAGAGCTACAAATTTAATGAAGGAACTGCATATGATGTAATACCAAGTA 663
DB 1000 GGAAGAGATTAACCTACTAAAGGAGCACTGGAACCTTTAGCAGATGAATAATCTCCTCT 1059
QY 664 ATATTAACTCTAAATCTACTAGTAAGATTAATTTAAAGACTGCAGTAGAAGAGTTACAA 723
DB 1060 AAAGTAACTAAAGAGCAGATAAAGTAAAGAACTTTAAAGATTTATGTTGAAGATTTAA 1119
QY 724 AAATTGAATGCTAGTTATCTTAACTACTACAACTTTAGCTGCTGATGACAGAAATACAAACA 783
DB 1120 AATGCTAACAAATGATATTCATATCTGTTGTTGTTAGCAGTGAAGATAGTAAGAAACA 1179
QY 784 GCTATAGAGATAAGTAAAGAAATATTAACAATAATGATGGCGAGAAATCAGATCAATTCAGCT 843
DB 1180 GCAATAGAGTTAAGTAGCAATACTATAAATCTGATGATGACAA-----TGCA 1227
QY 844 GATGTTAAAGAGAAATGTTAAAGATGTTGTTATAGTAGTGCGAAATGCACCTAGTAGATGGA 903
DB 1228 ATAACTAAAGATCCAGTTAACTGTTGTTTATGTTGTTTCTCAAGCTGATGTTGATGGG 1287
QY 904 TTAGTTCGGCTCTTTAGCAGCAAGAAAGATGCTCCACTATTTATTAATCTTCAAAAGAT 963
DB 1288 CTTGTAGCTTCACTTTAGCATCTGAAAGAAAGAGCTCTCTTACTATTAACTTACAGCAGGA 1347
QY 964 AAATTAGATGCTCAGTAAATCTGAAATAAGAGAGTTTTAGACTTTAAAGAACTTCAACA 1023
DB 1348 AAATTAGATGCTGTTAAAGCTGAGTTGAAAGAGTAAATGGATTTAAATCTTCAACA 1407
QY 1024 GAAGTA---ACAGGAAACAAAGTTTATATAGCTGGTGAGTTAAATAGTATCTAAAGAA 1080
DB 1408 GGTGTAATAACTTCTAAAGAAAGTTTACTTAGCTGGTGAGTAAACTCTATATCTAAAGAT 1467
QY 1081 GTTGTAAACAGATTAGATCAATCGGATTAAGCTTAAAGCTTGAAGATTTCTCAGGTGATGATGA 1140
DB 1468 GTAGAAATGAATTTAAAGATATGCGACTTAAAGTTTAAAGATTTATCAGGAGATGATGATGA 1527
QY 1141 TATGAAATCTTTTAAAGATAGCAGGTGAAATAGCTTTAGATATAATGATAGGCTTTATGTA 1200
DB 1528 TATGAAATCTTTTATGCTATGATGATGAAATAGCTTTGATATGATTAAGATCTTTTGTGA 1587

```

```

RESULT 5
US-10-239-610-4
; Sequence 4, Application US/10239610
; Publication No. US20040039165A1
; GENERAL INFORMATION:
; APPLICANT: Imperial College of Science Technology and Medicine
; TITLE OF INVENTION: CLOSTRIDIUM DIFFICILE POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: YOUZ 2 00089
; CURRENT FILING DATE: 2002-09-24
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: PCT/GB01/01305
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: GB 0007263.7
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2145
; TYPE: DNA
; ORGANISM: Clostridium difficile
US-10-239-610-4

Query Match      26.4%; Score 483.2; DB 18; Length 2145;
Best Local Similarity 62.7%; Pred. No. 2.6e-66;
Matches 911; Conservative 0; Mismatches 433; Indels 108; Gaps 6;

QY 394 ACTGTTGAAATTAAGTAGTAGCAGCTAGTGAAAGAAACAGTAGTTGTATCTAGTGATGCG 453

```

```

Db      784 ACTATAAAAGTTAGAGTTACAGTGCAAAAGAGAGATCTATTGATGTGATTCAGATTCA 843
QY      454 AAAAATAGTGCAAAAGATATAGCTGAAAAAATATGTGTTTGAAGACAAAGACTTTAGAAAT 513
Db      844 TATATTAGTGCTGAAATTTAGCTAAAAAATATGATTTAATCTCTAAAGAGGTTTCTGAA 903
QY      514 GCATTAAAGACTATAAATGCCCTCAGATTTTCAGTAAAGCTGATAGTTACTATCAAGTATGTT 573
Db      904 GCTTATAATAGCAATAGTTCGATTTACAAAATGATGGAATAGAAATCGATTTAGTACAATTA 963
QY      574 CTTTATCCAAAGGAAGAGATTACAAGGTTTCTCAACTTATAG---AGCTACAAATTAT 630
Db      964 GTTAATGGAATATCAAGTTATTTCTATCCAGAGGAAAGATTAGAACTAAATCT 1023
QY      631 AATGAAGAACTGCATATAGGTAAATACACAGTAATATTAATCTCTAAAATCTACTAGTAG 690
Db      1024 GCAGATATAATAGCTGATGCAGATAGTCCAGCTAAAAATAACTATATAAAGCTAAATAA 1083
QY      691 AGTAATTTAAAGACTGCAGTAGAAGATTACAAAATTCGAATTCGAATGCTTAATCTACT 750
Db      1084 AAGATTTTAAAGATTATGTAGATGATTTAAAAACATACAATAACTTACTCAATGTT 1143
QY      751 ACACTTTTAGCTGGTGATGACAGAAATACAAACAGCTTATAGAGATTAAGTAAAGAAATATTAC 810
Db      1144 GTACAGTAGCAGGAGAGATAGATAGAACTGCTATAGAAATTAAGTAGTAATATTAT 1203
QY      811 AATAATGATGGCAGAAATCAGATCAATCAGCTGATGTTTAAAGAGAAATGTTAAAAATGTT 870
Db      1204 AATCTGATGATAAAAAATGCAATACTGATGATGCAATGTTAAAT-----AATATA 1251
QY      871 GTATTAGTAGGTCGAATGCATAGTAGATGATGATTTGCGGCTCCTTTAGCAGCAGAA 930
Db      1252 GTATTAGTTGGATCTACATCTATAGTTGATGCTGTTGTCATACCATTAGCTTCAGAA 1311
QY      931 AAGATGCTCCACTATTTATTAATCTCAAAAGATAAATTAGATTGCTGAGTAAATCTGAA 990
Db      1312 AAAACAGCTCCATTATTTAACTTCAAGAGATAAATTAGATTTCATCAGTAAATCTGAG 1371
QY      991 ATAAGAGAGTTTTAGACTTAAATAAATCTCAACAGAGTA---ACAGGAAAAACAGTTTAT 1047
Db      1372 ATAAAAAGAGTTATGAATTTAAGAGTGATCTGGTATATAATCTCTTAAAAAAGTTTAT 1431
QY      1048 ATAGCTGGTGGATTAATAGTGTATCTTAAGAGTTGTTAAGAGATTAGAAATCAATGGA 1107
Db      1432 TTAGCTGGTGGATTAATCTATATCTAAGAGATGTAAGAAATGAATTTGAAAAATATGGGC 1491
QY      1108 TTAAGAGTTGAAAGATTCTCAGTGATGATAGATGAAATCTTTTAAAAATAGCAGGT 1167
Db      1492 CTTAAGATTACTAGATTATCAGGAGAGACAGATACGAATCTTTTAGCAATAGCTGAT 1551
QY      1168 GAAATAGGCTTAGATTAATAGAGCTTATAGTTGGTGGAAACAGGATTAGCAGATGCC 1227
Db      1552 GAAATAGGCTTGATATAAGCAATTTGATGTTGGTGGTACTGGATTAGCAGATGCT 1611
QY      1228 ATGAGTATAGCTTCACTGCTTCTACTAAATTAGATGGTAAATGGTGTAGATAGAAC 1287
Db      1612 ATGAGTATAGCTCCAGTTGCTCTCT-----CAACTTAAA 1644
QY      1288 AATGGACATGCTACTCCAATAGTTGTTGATAGTGAAGAGCTGATAAAAATATCTGATGAC 1347
Db      1645 GATGGAGATGCTACTCCAATAGTTGTTGATAGTGAAGAGCTGATAAAAATAGTATGAT 1704
QY      1348 TTAGATAGTTTCTTAGGAAGCGCTGATGATAGATATAATAGGTGGATTTTCAAGTGTATCT 1407
Db      1705 GCTAAGAGTTCTTAGGAAGCTTCTGATGTTGATATAATAGGTGGAAAAAATAGCGTATCT 1764
QY      1408 GAAAGATGGAAGAGCTATATCAGATGCTACTGTTAAAGCGGTACAGAGTTAAAGGC 1467
Db      1765 AAGAGATTGAAGAGTCAATAGATAGTGCACTGGAAAAAATCCAGATAGATAAAGTGA 1824
QY      1468 GACGATAGACAAGACTAACTCTGAAAGTTTAAAAAACATATATTATGCTTAATGATCTGAA 1527

```

```

Db      1825 GATGACAGACAAGCAACTAATGCTGAAGCTTTTAAAA----- 1860
QY      1528 ATAGCTAAAGCTGCAGTGTTTTAGATTAAGATTTCAGGTGCTTCAAGTAGTATGATGAGAGTA 1587
Db      1861 -----GAAGATGATTATTTTCAAGAGATGGTGAAGTT 1890
QY      1588 TTTAATTTCTATAGCTAAAGATGATCTACAAAAGAGATCAATTAGTTGATGCATTA 1647
Db      1891 GTGAATTAATTTGTTGCAAAAGATGTTCTACTAAAGAGATCAATTAGTAGATGATTA 1950
QY      1648 GCAGTAGGAGCTGTTGCTGG-----ATATAAACTTGCTCCAGTTGTATTAGCTACT 1698
Db      1951 GCAGCAGCACCAATAGCAGGTAGATTTAAGGAGTCTCCAGCTCCAATCATACTAGCTACT 2010
QY      1699 GATTCCTTTATCTTCTGATCAATCGGTTGCTATAGCAAAAGTTGTAGGAGAAAAATATTCT 1758
Db      2011 GATACCTTTATCTTCTGACCAAAATGATGCTGTAAAGTAAAGCAGTTCTTAAAGATGGTGA 2070
QY      1759 AAAGATTTTAAACACAAGTTGGTCAAGGAATAGCTAATTCAGTTTATATAACAAAATGAAAGAT 1818
Db      2071 ACTAACTTAGTTCAAGTAGGTAAAGGTATAGCTTCTTCAGTTTATAACAAAATGAAAGAT 2130
QY      1819 TTATTAGATATG 1830
Db      2131 TTATTAGATATG 2142

RESULT 6
US-10-068-870-7
; Sequence 7, Application US/10068870
; Publication No. US20030054009A1
; GENERAL INFORMATION:
; APPLICANT: The Provost, Fellows & Scholars of the College of the Holy and Undivided
; APPLICANT: Trinity of Queen Elizabeth, near Dublin
; TITLE OF INVENTION: C.difficile vaccine
; FILE REFERENCE: TRI002/C/WO
; CURRENT APPLICATION NUMBER: US/10/068,870
; CURRENT FILING DATE: 2002-02-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 2145
; TYPE: DNA
; ORGANISM: Clostridium difficile
US-10-068-870-7

Query Match      26.2%; Score 480; DB 14; Length 2145;
Best Local Similarity 62.6%; Pred. No. 8.4e-66;
Matches 909; Conservative 0; Mismatches 435; Indels 108; Gaps 6;

QY      394 ACTGTTGAAATTAAGTAGTAGCAGCTAGTGAAGAAACAGTAGTTGTATCTAGTATGCGG 453
Db      784 ACTATAAAAGTTAGAGTTACAGTGCAAAAGAGAGAAATCTATTGATGTGGAATTCAGATTCA 843
QY      454 AAAAATAGTGCAAAAGATATAGCTGAAAAAATATGTGTTTGAAGACAAAGACTTTAGAAAT 513
Db      844 TATATTAGTGTGAAAAATTTAGCTAAAAAATATGTTATTATCTCTTAAAGAGGTTTCTGAA 903
QY      514 GCACCTAAAAAACTATAAATGCTCTCAGATTTCAGTAAAACTGATAGTTACTATCAAGTAGTT 573
Db      904 GCTTATAATGCAATAGTTGCAATTACAAAATGATGGAATAGAAATCTGATTAGTACAATTA 963
QY      574 CTTTATCCAAAGGAAGAGATTAAGAGTTTCTCAACTTATAG---AGCTACAAATTA 630
Db      964 GTTAATGAAAAATATCAAGTTATTTTCTATCCAGAGGAAAAAGATTAGAACTAAATCT 1023
QY      631 AATGAAGAACTGCATATGTTAATACACAGTAATATTAACTCTAAAATCTACTAGTAG 690
Db      1024 GCAGATATAATAGCTGATGCGATAGTCCAGCTAAATTAATCTATAAAGCTTAATAATTA 1083
QY      691 AGTAATTTAAAGACTGCAGTAGAAGAGTTACAAAATTTGAATGCTAGTTATTCTAATACT 750
Db      1084 AAGATTTTAAAGATTATGTAGATGATTTAAAAACATACAATAATATTACTTACTCAATGTT 1143

```

QY 751 ACACTTTAGCTGGTATGACAGAAATACAAACAGCTATATAGAGATAAGTAAGAAATATTAC 810
 Db |||||
 QY 1144 GTAAACAGTAGCAGGAGAGATAGAAATGCTATAGAAATTAAGTAGTAAATATTAT 1203
 Db |||||
 QY 811 AATAATGATGGCAGAGAAATCAGATCAATTCAGCTGATGTTAAAGAGAAATGTTAAAAATGTT 870
 Db |||||
 QY 1204 AATCTGATGATAAATGCAATCACTGATGATGCGAGTTAAT-----AATATA 1251
 Db |||||
 QY 871 GTATTAGTAGGTCAGAAATGCACTAGTAGATGGAATTTAGTTCGGCTCTTTAGCAGAGAA 930
 Db |||||
 QY 1252 GTATTAGTAGGTCATCATCTATAGTTGATGGTCTTTGTCATCACCATTAGCTTCAGAA 1311
 Db |||||
 QY 931 AAGATGCTCCATTTATTAATCTCAAAAGATAAATTTAGATTCGTCAGTAAATCTGAA 990
 Db |||||
 QY 1312 AAAACAGCTCCATTTATTAATTAATCTCAAAAGATAAATTTAGATTCATCAGTAAATCTGAG 1371
 Db |||||
 QY 991 ATAAGAGAGCTTTTAGACTTTAAAACTTTCAACAGAGTA---ACAGGAAAAACAGTTTAT 1047
 Db |||||
 QY 1372 ATAAAAGAGTTATGAACTTTAAGAGTGATCTGGTATAAATACTTCTAAAAAAGTTTAT 1431
 Db |||||
 QY 1048 ATAGCTGGTAGGTTAATAGTGTATCTAAAGAGTTGTTAAGAGAAATTTAGAAATCAATGGGA 1107
 Db |||||
 QY 1432 TTAGCTGGTAGGTTAATTTCTATATCTAAAGATGTTAGAGATGAATTTGAAAAATATGGGC 1491
 Db |||||
 QY 1108 TTAAGAGTTGAAAGATTTCTCAGGTGATGATAGATATGAAATCTTTTAAAAATAGCAGGT 1167
 Db |||||
 QY 1492 CTTAAGGTTTACTAGATTTATCAGGAGAGAGAGATACGAAATCTTTTAGCAATAGCTGAT 1551
 Db |||||
 QY 1168 GAAATAGGCTTTAGATAATGATAGGCTTTATGTTAGTTCGGTGAACAGAGATTTAGCAGATGCC 1227
 Db |||||
 QY 1552 GAAATAGGCTTTGATATGATTAAGCAATTTGTTAGTTCGGTGTCTGGATTTGCGCAGATGCT 1611
 Db |||||
 QY 1228 ATGAGTATAGCTTCAGTTGCTTCTTCTAATAATTTAGATCGTAATGTTGTTAGATAGAAACA 1287
 Db |||||
 QY 1612 ATGAGTATAGCTTCAGTTGCTTCTCT-----CAACTTAAA 1644
 Db |||||
 QY 1288 AATGACATGCTACTCTCAATAGTTGTTGATAGTGAAGCTGATGAATATCTGATGAC 1347
 Db |||||
 QY 1645 GATGAGATGCTACTCTCAATAGTATGTTGATAGTGAAGCAAGAAATTAAGTATGAT 1704
 Db |||||
 QY 1348 TTAGATAGTTTCTTAGGAGCGCTGATGATAGATATAATAGTGATTTGCAAGTGTATCT 1407
 Db |||||
 QY 1705 GCTAAGAGTTCTTAGGAACTTCTGATGTTGATATAATAGTGGAAATATAGCTATCT 1764
 Db |||||
 QY 1408 GAAAGAGTGAAGAGCTATATCAGATGCTACTCGTTAAGCGGTTAACAAGTTAAAGGC 1467
 Db |||||
 QY 1765 AAAGAGATTGAAGAGTCAATAGATAGTGCAACTCGAAAACTCCAGATAGAATAAGTGA 1824
 Db |||||
 QY 1468 GACATAGACAAAGACACTTAATCTGAAAGTTATATAAACAATATTATGCTAATGACTGAA 1527
 Db |||||
 QY 1825 GATGACAGACAAAGCACTTAATGCTGAAAGTTTAAAA----- 1860
 Db |||||
 QY 1528 ATAGCTAAAGCTGCAGTTTTAGATAAAGATTCAGGTGCTTCAAGTAGTGATGACAGAGTA 1587
 Db |||||
 QY 1861 -----GAAATGATTAATTTCAAGATGGTGAAGTT 1890
 Db |||||
 QY 1588 TTTAATTTCTATGCTAGCTAAAGATGGAATCTCAAAAGAGATCAATTTAGTTGATGCAATTA 1647
 Db |||||
 QY 1891 GTGAATTTCTTTGTTGCAAAAGATGGTTCTACTAAAGAGATCAATTTAGTAGATGCAATTA 1950
 Db |||||
 QY 1648 GCAGTAGAGCTGTTGCTGG-----ATATAAACTTGGCTCCAGTTGTTATTAGCTACT 1698
 Db |||||
 QY 1951 GCAGCAGCACCAGTAAAGAGTAGATTTAAGGAGTCTCCAGCTTCAATCATACTAGCTACT 2010
 Db |||||
 QY 1699 GATCTTTTATCTTCTGATCAATCGGTCTGATAGCAAGTTGTTAGGAGAAAAATATTCT 1758
 Db |||||
 QY 2011 GATACCTTTATCTTCTGACCAAAATGATGCTGTTAAGTAAAGCAGTTCTTAAAGATGGTGA 2070
 Db |||||
 QY 1759 AAAGATTTTAAACAAGTTGGTCAAGGAATAGCTAATTCAGTTTATAACAATAAGATGAAGAT 1818
 Db |||||
 QY 2071 ACTAATCTTAGTTCAAGTAGGTAAGGATAGCTTCTTCCAGTTTATAACAATAAGATGAAGAT 2130
 Db |||||

QY 1819 TTATTAGATATG 1830
 Db |||||
 QY 2131 TTATTAGATATG 2142
 Db |||||
 RESULT 7
 US-10-068-870-1
 ; Sequence 1, Application US/10068870
 ; Publication No. US20030054009A1
 ; GENERAL INFORMATION:
 ; APPLICANT: The Provost, Fellows & Scholars of the College of the Holy and Undivided
 ; TITLE OF INVENTION: Trinity of Queen Elizabeth, near Dublin
 ; FILE REFERENCE: TRI002/C/WO
 ; CURRENT APPLICATION NUMBER: US/10/068,870
 ; CURRENT FILING DATE: 2002-02-11
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 2157
 ; TYPE: DNA
 ; ORGANISM: Clostridium difficile
 ; US-10-068-870-1
 Query Match 25.8%; Score 471.8; DB 14; Length 2157;
 Best Local Similarity 61.5%; Pred. No. 1.6e-64;
 Matches 951; Conservative 0; Mismatches 467; Indels 129; Gaps 7;
 QY 296 CTACAGCTAATGGAATGAAGATTATGTAAGACAACTTTAAAAATTTAGATGCGAGGAG 355
 Db |||||
 QY 728 CAACACCTAGTCTGCTAGCTGTAAGTGGTTTGTAACTAAAGATGATCTGATTTAGCAA 787
 Db |||||
 QY 356 AATATCTCTATTATAGATTTAACTTATAATAATGCTAAAACTGTTGAAATTTAAAGTAGTAG 415
 Db |||||
 QY 788 AATCAGTACTATTAATGTAAGGTTTAAATGCAAAAGAGATCAATGATATAGATG 847
 Db |||||
 QY 416 CAGCTAGTGAATAAAGACAGTAGTTGTTATCTAGTATGCGAAAAATAGTGCAAAAAGATATAG 475
 Db |||||
 QY 848 CAGCTCATATATATCATCAGCTG-----AAAAATTAGCTAAAAAGATATGTTATTTGATCCAG 901
 Db |||||
 QY 476 CTGAAAAATATGTTGTTGAAGACAAAGACTTTAGAAAATGCACCTAAAAAATATATAATGSCCT 535
 Db |||||
 QY 902 ATGAAATTTCTGAAGCATATAGGCAATAGTAGCATTTACAAAATGATGGTATAGATCTA 961
 Db |||||
 QY 536 CAGATTTCACTAAAACTGATAGTTACTATCAAGTAGTCTTTATCCAAAAAGGAAAGAGAT 595
 Db |||||
 QY 962 ATTTAGTTCAGTTAGTTAATGGAATATCAAGTATTTTATCCAGAGGTAAGAGAT 1021
 Db |||||
 QY 596 TACAAGGTTTTCTCAACTTATAGAGCTACAAATTTAATGAAGGAACCTGCATATGGTAATA 655
 Db |||||
 QY 1022 TAGAACTAAATCAGCA-----AATGATACAATAGCTAGTCAAGATA 1063
 Db |||||
 QY 656 CACAGTAAATTTAACTCTTAAATCTACTAGTAAGAGTAATTTAAAGACTGCGAGTAGAAG 715
 Db |||||
 QY 1064 CACCAGCTAAGTAGTTATATAAGCTAATAAATTTAAAGATTTAAAAAGATTTATGTAGATG 1123
 Db |||||
 QY 716 AGTTACAAAAATTTGAATGCTAGTTATTCTAATCTACTACAACTTTAGCTGGTGCATGACAGAA 775
 Db |||||
 QY 1124 ATTTAAAAACATATATATATCTTTTCAATGTTGTTAAACAGTAGCAGGAGAGATAGAA 1183
 Db |||||
 QY 776 TACAAAACAGCTATAGAGATAAGTAAGAAATATTACAATAATGATGGCGAGAAATCAGATC 835
 Db |||||
 QY 1184 TAGAACTGCTATAGAATTAAGTAGTAAATATTATATATCTCGAT-----GATA 1231
 Db |||||
 QY 836 ATTCAGCTGATGTTAAAGAGAAATGTTAAAAATGTTGTTATTAGTAGGTGCAATGCACATAG 895
 Db |||||
 QY 1232 AAAATGCAATAACTGATATAAGCAGTTAATGATATAGTATTAGTTGGATCTACATCTATAG 1291
 Db |||||
 QY 896 TAGATGATTTAGTTGGCTCTCTTTAGCAGCAGAAAAAGATGCTCCACTATTATTAACTT 955
 Db |||||
 QY 1292 TTGATGCTGTTGTTGATCACTACCTAGCTTCCAGAAAAAACAGCTCCATTTATTAACTT 1351
 Db |||||
 QY 956 CAAAAGATAAATTAGATTTCGTTCAGTAAAAATCTGAAAAATGAAGAGAGTTTTAGACTTAAAAA 1015
 Db |||||

```

1352 CAAAAGTAAATAGATTATCATAGTAAATCTGAAATAAAGAGATTGAACTTAAAGA 1411
1016 CTTCAACAGAGTA---ACAGGAAAAACAGTTTATATAGCTGCTGGTGAATTAATAGTAT 1072
1412 GTGACACTGGTATAAATACTTCTAAAAAAGTTTATTTAGCTGGTGAATTAATCTATAT 1471
1073 CTAAGAAGTTGTAAACAGATTAGATCAATGGGATTAAGGTTAAAGATTTCAAGGTG 1132
1472 CTAAGATGTAGAAAAATGAATGAAAAACATGGGCTTTAAAGTTACTAGATTATCAGGAG 1531
1133 ATGATAGATATGAACCTTCTTTAAAAATAGCAGTGAATAGGCTTAGATTAATGATAAGG 1192
1532 AAGACAGATACGAACCTTCTTTAGCAATAGCTGATGAAATAGTCTTGATTAATGATAAG 1591
1193 CTTATAGTTGGTGAACAGGATTAGCAGATGCCATGATAGTATAGCTTCAGTTGCTTCTA 1252
1592 CATTTGTAGTTGGTGGTACTGGATTAGCAGATGCTATGATAGTATAGCTCCAGTTGCTTCT- 1650
1253 CTAATATAGTGTATGTTGTAGATAGAACAAATGGACATGCTACTCCATATAGTTG 1312
1651 -----CAAACCTTAAGATGGAGATGCTACTCCAAATAGTAG 1684
1313 TTGTAGATGAAAAGCTGATAAATATCTGATGACTTAGATAGTTTCTTAGGAAGCGCTG 1372
1685 TTGTAGATGGAAGAAGCAAGAATAAGTGATGATGCTGAAGTTTCTTAGGAACCTTCTG 1744
1373 ATGTAGATATAATAGGTGGATTTGCAAGTGTATCTGAAAAGATGGAAGAGCTTATATCAG 1432
1745 ATGTTGATATAATAGGTGGAAAAAATAGCGTATCTAAAGAGATTGAAGAGTCAATAGATA 1804
1433 ATGCTACTGGTAAGCGGTTACAAGAGTTAAAGGCGACGATAGACAAAGACATACTCTG 1492
1805 GTGCAACTGGAAAAAATCTCAGATAGAAATAAGTGGAGATGATAGACAAGCAACTAATGCTG 1864
1493 AAGTTATAAAAAACATATTATGCTTAATGATACTGAAATAGCTAAAGCTGCAGTTTTAGATA 1552
1865 AAGTTTAAAGAAAGATGATTATTC-----1890
1553 AAGATTCAAGTGCTTCAAGTAGTGATGCGAGGAGTATTTAATTTCTATGCTAGCTAAAGATG 1612
1891 -----ACAGATGGTGAAGTTGTGAATTAATTTGTTGCAAAAAGATG 1930
1613 GATCTACAAAAGAGATCAATTAGTTGATGCGATTAGCAGTACGAGCTGTTGCTGG-----1667
1931 GTTCTACTAAGAAGATCAATTAGTAGATGCGCTTAGCAGCGACCAATAGCAGGTAGAT 1990
1668 ----ATATAAAGCTTGCCTCAGTTGTTATAGCTACTGATTTCTTTATCTCTGATCAATCGG 1723
1991 TTAAAGAGTCTCAGCTCCATCATATCTAGCTACTGATTTATTTCTCTGACCAAAATG 2050
1724 TTGCTATAAGCAAGTTGTAGGAGAAAAATATTTCTAAAGATTTTAAACAAAGTTGGTCAAG 1783
2051 TAGCTGTAAGTAAAGCAGTTCTCTAAAGATGGTGGAACTAACTTAGTTTCAAGTAGGTAAAG 2110
1784 GAATAGCTAATTCAGTTTAAACAAATGAAAGATTTATAGTATG 1830
2111 GTATAGCTTCTTCAGTTTAAACAAATGAAAGATTTATTAGATATG 2157

```

```

RESULT 8
US-10-068-870-3
; Sequence 3, Application US/10068870
; Publication No. US20030054009A1
; GENERAL INFORMATION:
; APPLICANT: The Provost, Fellows & Scholars of the College of the Holy and Undivided
; TITLE OF INVENTION: C difficile vaccine
; FILE REFERENCE: TRI002/C/WO
; CURRENT APPLICATION NUMBER: US/10/068, 870
; CURRENT FILING DATE: 2002-02-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1

```

```

; SEQ ID NO 3
; LENGTH: 2158
; TYPE: DNA
; ORGANISM: Clostridium difficile
US-10-068-870-3

Query Match      25.8%; Score 471.8; DB 14; Length 2158;
Best Local Similarity 61.5%; Pred. No. 1.6e-64;
Matches 951; Conservative 0; Mismatches 467; Indels 129; Gaps 7;

QY 296 CTCAGCTAAATGAAGATTATGTAAAGACAACTTTTAAAAATTTAGATGCGAGAG 355
DB 728 CAACACCTAGTCTGTAGCTGTAAGTGGTTTGTAACTAAAGATGATCTAGTTTAGCAA 787
QY 356 AATATGCTATTATAGATTTAACCTTATATAATGCTTAAAACTGTTGAAATTAAGTAGTAG 415
DB 788 AATCAGGTACTATAAATGTAAAGATTATAAATGCAAAAGAAAGATCAATTTGATATAGATG 847
QY 416 CAGCTAGTGAAAAACAGTAGTTGTATCTAGTGATGCGAAAAAATAGTGCAAAAGATATAG 475
DB 848 CAAAGCTCATATACATCAGCTG-----AAAATTTAGCTAAAGATATGTAATTTGATCCAG 901
QY 476 CTGAAAAATATATGTTTGAAGACAAAGACTTAGAAAAATGCATAAAAACTATAAATGCCCT 535
DB 902 ATGAAATTTCTGAAGCATATAGGCAATAGTAGCATTACAAATGATGCTATAGAGTCTTA 961
QY 536 CAGATTTCAAGTAAAACTGATAGTTACTATCAAGTAGTTCTTTTATCCAAAAGGAAAGAGAT 595
DB 962 ACTTAGTTTCAGTTAGTTAATGGAATAATATCAAGTGTATTTTATCCAGAAGGTAAAGAT 1021
QY 596 TACAAGGTTTCTCAACTTATAGAGCTACAAAATATATAATGAAGAACTGCATATGTTAATA 655
DB 1022 TAGAACTAAATCAGCA-----AATGATACAATAGTAGTCAAGATA 1063
QY 656 CACCAGTAAATATAACTCTTAAATCTACTAGTAAGAGTAAATTTAAAGACTGCGAGTAGAG 715
DB 1064 CACCAGCTAAAGTAGTTATAAAGCTAATAAATTTAAAGATTTTAAAGATTTATGTAGATG 1123
QY 716 AGTTACAAAATTTGAATGCTAGTTATTTCTAATCTACAACTTTAGCTGGTGTATGACAGAA 775
DB 1124 ATTTAAAAACATATATAATTAATCTTCAATGTTTGAACAGTACGAGGAGAGATAGAA 1183
QY 776 TACAACAGCTATAGAGATAAGTAAAGAAATATTACAATAATGATGCGGAGAAATCAGATC 835
DB 1184 TAGAACTGCTATAGAAATTAAGTAGTAATAATTTAATTTCTGAT-----GATA 1231
QY 836 ATTCAGCTGATGTTAAAGAGAAATGTTTAAAAATGTTGTATTTAGTAGTGCAATGCACTAG 895
DB 1232 AAAATGCAATAACTGATAAGCAGTTAATGATATAGTATTAGTTGGATCTACATCTATAG 1291
QY 896 TAGATGGATTTAGTTGGGCTCCTTTAGCAGCAGAAAAAGATGCTCCACTATTATTAACCTT 955
DB 1292 TTGATGGTCTTGTGTCATCACCATTAGCTTCAGAAAAAACAAGCTCCACTTATTATTAACCT 1351
QY 956 CAAAAGATAAATTTAGATTGCTCAGTAAATCTGAAATAAAGAGAGTTTTTGTAGACTTTAAAA 1015
DB 1352 CAAAAGATAAATTTAGATTTCATCAGTAAATCTGAAATAAAGAGAGTTTGAATCTTAAGA 1411
QY 1016 CTTCAACAGAGAAGTA---ACAGGAAAAACAGTTTTATATAGCTGGTGGAGTTAATAGTGTAT 1072
DB 1412 GTGACACTGGTATAAATCTTCTTAAAAAAGTTTATTTAGCTGGTGGAGTTAATTTCTATAT 1471
QY 1073 CTAAGAAGATTGTATAACAGAAATTTAGAAATCAATGGGATTTAAAGTTGAAAGATTTCTCAGGTG 1132
DB 1472 CTAAGAAGATTGTAGAAAAATGAATTTGAAAAAACAATGGGCTTTAAAGTTACTAGATTATCAGGAG 1531
QY 1133 ATGATAGATATCAAACTTCTTTAAAAATAGCAGGTGAAATAGGCTTAGATAATGATAAGG 1192
DB 1532 AAGACAGATACGAACCTTCTTTAGCAATAGCTGATGAAATAGTCTTTGATTAATGATAAG 1591
QY 1193 CTTATGATGTTGGTGAACAGGATTTAGCAGATGCCATGATAGTATAGCTTCAGTTGCTTCTTA 1252
DB 1592 CATTTGTAGTTGGTGGTACTGGATTAGCAGATGCTATGATAGTATAGCTCCAGTTGCTTCT- 1650

```

QY 1253 CTAAATTAGATGGTAATGGTGTGTAGATAGAACAAATGACATGCTACTCCCAATAGTTG 1312
 Db 1651 -----CAACTTAAAGATGGATGCTACTCCCAATAGTAG 1684
 QY 1313 TTGTAGATGGAAGCTGATAAAATATCTGATGACTTTAGATAGTTTCTTAGGAAGCGCTG 1372
 Db 1685 TTGTAGATGGAAGCAAAAGAAATAGTGTATGATGCTAAGAGTTTCTTAGGAACCTCTG 1744
 QY 1373 ATGTAGATATAATAGTGGATTTGCAAGTGTATCTGAAAGATGGAAGAGCTTATATCAG 1432
 Db 1745 ATGTTGATATAATAGTGGGAAAAATAGCGTATCTTAAAGAGATTTGAAGAGCTCAATAGATA 1804
 QY 1433 ATGCTACTGTAAAGGGGTTTACAAGAGTTTAAAGCGACGATAGACAAAGACACTAACTCTG 1492
 Db 1805 GTGCAACTGGAAAAAAGCTCCAGATAGATAATAGTGGAGATGATAGCAAGCAACTATAGCTG 1864
 QY 1493 AAGTTATAAAAAACATATATGCTAATGATATCTGAAATAGCTAAAGCTGCAGTTTTAGATA 1552
 Db 1865 AAGTTTAAAGAGATGATATTTTC----- 1890
 QY 1553 AAGATTCAGGTGCTTCAAGTAGTATGATGACAGAGATTTAAATTTCTATGTAGCTAAAGATG 1612
 Db 1891 -----ACAGATGCTGAAGTTGTGAATTAATCTTTGTTGCAAAAGATG 1930
 QY 1613 GATCTACAAGAGATCAATAGTTCGATGCTATGATGCTAGTGGAGTGTGCTGG----- 1667
 Db 1931 GTTCTACTAAGAAGAGATCAATTAGTAGTGCCTTAGCAGCAGCACCNAATAGCAGGTAGAT 1990
 QY 1668 ----ATATAAACTTGCTCCAGTTCGTATTTAGTACTGATTCCTTTATCTTCTGATCAATCGG 1723
 Db 1991 TTAAGGAGTCTCCAGTCTCAATCATACTAGTACTGATCTTTATCTTCTGACCAAAATG 2050
 QY 1724 TTGCTATAAGCAAAAGTTGTAGGAGAAATATCTTAAAGATTTAAACACAGTTGGTCAAG 1783
 Db 2051 TAGCTGAAGTAAAGCAGTTCCTTAAAGATGTTGGTGAACCTAACTTAGTTTCAAGTAGGTAAAG 2110
 QY 1784 GAATAGCTAATTCAGTTATTAACAAATGAAGATTTTATTAGATATG 1830
 Db 2111 GTATAGCTTCTTCAAGTTATTAACAAATGAAGATTTTATTAGATATG 2157

RESULT 9

; US-10-068-870-8
 ; Sequence 8, Application US/10068870
 ; Publication No. US20030054009A1
 ; GENERAL INFORMATION:
 ; APPLICANT: The Provost, Fellows & Scholars of the College of the Holy and Undivided
 ; APPLICANT: Trinity of Queen Elizabeth, near Dublin
 ; TITLE OF INVENTION: C.difficile vaccine
 ; FILE REFERENCE: TRI002/C/WO
 ; CURRENT APPLICATION NUMBER: US/10/068,870
 ; CURRENT FILING DATE: 2002-02-11
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 8
 ; LENGTH: 2158
 ; TYPE: DNA
 ; ORGANISM: Clostridium difficile
 ; US-10-068-870-8

Query Match 25.8%; Score 471.8; DB 14; Length 2158;
 Best Local Similarity 61.5%; Pred. No. 1.6e-64;
 Matches 951; Conservative 0; Mismatches 467; Indels 129; Gaps 7;
 QY 296 CTACAGCTAATGGAATGAAGATTATGTAAGACAACTTTTAAAAATTTTAGATGCGAG 355
 Db 728 CAACACCTAGTCTGTAGCTGTAAGTGGTTTTTGTAACTTAAAGATGATGATGATTTAGCAA 787
 QY 356 AATATGCTATTATAGATTAACTTATAATTAATGCTTAAACCTGTTGAAATTTAAAGTAGTAG 415
 Db 788 AATCAGGTACTATAATGTAAGAGTTATAAATGCAAAAGAAAGAAATCAATTTGATATAGATG 847

QY 416 CAGCTAGTGAAGAAAAACAGTAGTTGTATCTAGTGTATCGGAAAAAATAGTGCAGAAAGATATAG 475
 Db 848 CAAGCTCATATACATCAGCTG-----AAAAATTAGCTAAAGACATGTTATTGTGATCCAG 901
 QY 476 CTGAAAAATATGTTGTTGAAGACAAAGACTTTAGAAATGCACCTAAAAACCTATAAAATGCCT 535
 Db 902 ATGAAATTTCTGAAGCATATATAGGCATATAGTCATTACAAATGATGGTATATAGATCTTA 961
 QY 536 CAGATTTTCAGTAAAACTGATAGTTACTATCAAGTAGTCTTCTTATCCAAAGGAAAGAGAT 595
 Db 962 ATTTAGTTCAGTTAGTTAATGGAAAAATATCAAGTGTATTTTATCCAGAGGTTAAAGAT 1021
 QY 596 TACAAGGTTTCTCAACTTATAGAGCTCAAAATTTATTAATGAAGAACTGCATATGGTAATA 655
 Db 1022 TAGAAACTAAATCAGCA-----AATGATACAAATAGCTAGTCAAGATA 1063
 QY 656 CACCACTAATATTAACCTCTAAATCTACTAGTAAGAGTAAATTTAAAGAGCTGCAGTAGTAAG 715
 Db 1064 CACCACTAAGTAGTTATAAAAGCTAATAAATTAAGATTTTAAAGATTTATGTTAGATG 1123
 QY 716 AGTTACAAAAATTTGAATGCTAGTTATTTCTAATACTACAACCTTTAGCTGGTGTGACAGAA 775
 Db 1124 ATTTAAACATATATATACTTTTCAAAATGTTGTAAACAGTAGCAGGAGAAAGATAGAA 1183
 QY 776 TACAAACAGCTATAGAGATAAGTAAAGATATTTACAAATATATGATGGCGAGAAATCAGATC 835
 Db 1184 TAGAACTGCTATAGAAATTAAGTAGTAAATTTATTAATTTCTGAT-----GATA 1231
 QY 836 ATTCAGCTGATGTTAAAGAGAAATGTTAAAAATCTGTTATTAGTAGTGGTGCACATGCACATG 895
 Db 1232 AAAATGCAATPAACCTGATTAAGCAGTTAATGATATAGTATTAGTTGGATCTACATCTATAG 1291
 QY 896 TAGATGATAGTTGTTGGGCTCTTTTAGCAGCAGAAAAAGATGCTCCACTATTATTAATTA 955
 Db 1292 TTGATGCTCTGTTGGATCACCATTAGCTTCAGAAAAACAGCTCCATTTATTTAACTT 1351
 QY 956 CAAAGAGATAAATAGATTCCTCAGTAAAAATCTGAAATAAAGAGAGTTTATAGACTTAAAA 1015
 Db 1352 CAAAGAGATAAATAGATTCATCACTAGTAAAAATCTGAAATAAAGAGAGTTTATGAACCTTAA 1411
 QY 1016 CTTCAACAGAGTA--ACAGGAAAAACAGTTTATATAGCTGTTGGTGAAGTTAAATAGTGTAT 1072
 Db 1412 GTGACACTGGTATAAATACTTCTAAAAAAGTTTATTTAGCTGGTGGAGTTAATTTCTATAT 1471
 QY 1073 CTAAGAAAGTTGTAACAGAAATTAGAATCAATGGGATTTAAAGTTTGAAGATTTCTCAGGTG 1132
 Db 1472 CTAAGATGTAGAAAAATGAATTTGAAAAACATGGCTTTAAAGTTTACTAGATTATCAGGAG 1531
 QY 1133 ATGATAGATATGAACCTTCTTTAAAAATAGCAGGTGAATAGGCTTTAGATTAATGATTAAGG 1192
 Db 1532 AAGACAGATACGAAACTTCTTTAGCAATAGCTGATGAATAGGCTCTTTGATTAATGATTAAG 1591
 QY 1193 CTTATGATGTTGGTGGAAACAGGATTTAGCAGATGCCATGATGATAGCTTCACTGCTCTTCTA 1252
 Db 1592 CATTTGATGTTGGTGGTACTGGATTAGCAGATGCTATGATGATAGCTCCAGTTGCTTCT- 1650
 QY 1253 CTAAATTAGTGGTAAATGGTGTGTAGATAGAAACAAATGGACATGCTACTCCAATAGTTG 1312
 Db 1651 -----CAACTTAAAGATGGAGATGCTACTCCAATAGTAG 1684
 QY 1313 TTGTAGATGGAAGAGCTGATAAAATATCTGATGACTTAGATAGTTTCTTAGGAAGCGCTG 1372
 Db 1685 TTGTAGATGGAAGAGCAAAAGAAATAGTGTATGATGCTTAAGAGTTTCTTAGGAACCTTCTG 1744
 QY 1373 ATGTAGATATATAGTGGTGGATTTTCAAGTGTATCTGAAAGATGGAAGAGCTATATCAG 1432
 Db 1745 ATGTTGATATAATAGTGGAAAAAATAGCGTATCTTAAAGAGATTTGAAGAGCTCAATAGATA 1804
 QY 1433 ATGCTACTGTTAAAGGGCTTTACAAGAGTTTAAAGCGACGATAGACAAAGACACTAACTCTG 1492
 Db 1805 GTGCAACTGGAAAAAAGCTCCAGATAGATAATAGTGGAGATGATAGACACAGCAACTAATGCTG 1864
 QY 1493 AAGTTATAAAAAACATATTTATGCTAATGATCTGAAATAGCTTAAAGCTGAGCTGATTTAGATA 1552

```
Db 1865 AGTTTTAAAGAGATGATTTC----- 1890
Qy 1553 AAGATTGAGTCTTCAAGTAGTGATGCGAGGAGTATTTAAATTTCTATGTAGCTAAAGATG 1612
Db 1891 -----ACAGATGGTGAAGTGTGGAATTACTTTTGTGCAAAAGATG 1930
Qy 1613 GATCTACAAAGAGATCAATTAGTTAGTTCATGAGCATTAGCAGTAGGAGCTGTTGCTGG----- 1667
Db 1931 GTTCTACTAAGAGAGATCAATTAGTAGATGCCCTTAGCAGCAGCACCAATAGCAGGTAGAT 1990
Qy 1668 ----ATATAAACTTGGCTCCAGTTGTTATAGTCTACTGATCTTCTTATCTTCTGATCAATCGS 1723
Db 1991 TTAAGAGATCTCAGCTCCCAATCATACTAGTCTACTGATCTTTATCTTCTGACCAAAATG 2050
Qy 1724 TTGCTATAAGCAAGTTGTAGGAGAAAAATATTTCTAAAGATTTAAACAAGTTGGTCAAG 1783
Db 2051 TAGCTGTAAGTAAGCAGTTCTTAAAGATGGTGGAACTAACTTTAGTTCAAGTAGGTAAG 2110
Qy 1784 GAATAGCTAATTGAGTTATTAACAAAATGAAGATTTTATAGATATG 1830
Db 2111 GTATAGCTTCTTCAGTTATAAACAATAAGAAATGAAAGATTTTATAGATATG 2157

RESULT 10
US-10-239-610-5
; Sequence 5, Application US/10239610
; Publication No. US20040039165A1
; GENERAL INFORMATION:
; APPLICANT: Imperial College of Science Technology and Medicine
; TITLE OF INVENTION: CLOSTRIDIUM DIFFICILE POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: YOUZ 2 00089
; CURRENT APPLICATION NUMBER: US/10/239, 610
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/GB01/01305
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: GB 0007263.7
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2160
; TYPE: DNA
; ORGANISM: Clostridium difficile
US-10-239-610-5

Query Match 25.8%; Score 471.8; DB 18; Length 2160;
Best Local Similarity 61.5%; Pred. No. 1.6e-64;
Matches 951; Conservative 0; Mismatches 467; Indels 129; Gaps 7;

Qy 296 CTACAGCTAATGGAATGAAGATTATGTAAAGCAACTTTAAAGATTTAGATGCGAGGAG 355
Db 728 CAACAGCTAGTCTGTAGCTGTAAGTGGTTTGTAACTAAAGATGATGATCTGATTTAGCAA 787
Qy 356 AATATGCTATTATAGATTAACTTTATAATGCTAAAGCTTTGAAATTTAAAGTAGTAG 415
Db 788 AATCAGGTACTATAATGTAAGAGTTATAAATGCAAAAGAAAGAAATCAATTTGATATAGATG 847
Qy 416 CAGCTAGTGAAGAAACAGTAGTTGTATCTAGTATGCGAAAAATAGTGCAAAAGATATAG 475
Db 848 CAAGCTCATATACATCAGCTG-----AAAATTTAGCTAAAGATATGATTTGTATCCAG 901
Qy 476 CTGAAAAATATGTTGTTGAAGACAAAGACTTAGAAAAATGCACTAAAGATTTATTAATGCTT 535
Db 902 ATGAAATTTCTGAAGCATATATAGGCAATAGTACATTACAAAATGATGGTATAGAGTCTA 961
Qy 536 CAGATTTCAGTAAAACTGATAGTTACTATCAAGTAGTCTTTTATCCAAAGGAAGAGAT 595
Db 962 ATTTAGTTCAGTTAGTTAATGGAATAATATCAAGTGATTTTTTATCCAGAAAGGTAAAGAT 1021
Qy 596 TACAAGGTTTCTCAACTTATAGAGCTACAAATTTATTAATGAAGGAACTGCTATATGTAATA 655
Db 1022 TAGAACTAAATCAGCA-----AATGATACAATAGCTAGTCAAGATA 1063
```

```
Qy 656 CACCAGTAATATTAATCTTAAATCTACTAGTAAGAGTAATTTTAAAGACTGCGAGTAGAAG 715
Db 1064 CACCAGCTAAAGTAGTTTAAAGCTTAAATTTAAAGATTTAAAGAGTTATGTAGATG 1123
Qy 716 AGTTACAAAATTTGAATGCTAGTATTTCTAACTACTAACATTTTACGTTAGCTGGTATGACAGAA 775
Db 1124 ATTTAAACAATATTAATATCTTATTTCAATGTTGTTAACAGTAGCAGGAGGAAGATAGAA 1183
Qy 776 TACAACAGCTATAGAGATAAGTAAAGAAATATTACAATAATGATGGCGAGAAATCAGATC 835
Db 1184 TAGAACTGCTATAGATTAAGTAGTAAATATTATAATTTCTGAT-----GATA 1231
Qy 836 ATTGAGCTGATGTTTAAAGAGAAATGTTAAAAATGTTGTTATTTAGTAGTGCAGAAATGCACTAG 895
Db 1232 AAAATGCAATAACTGATAAAGCAGTTAATGATATAGTATTTAGTTGGATCTCATCTCATAG 1291
Qy 896 TAGATGGATTAGTTGCGGCTCCTTTAGCAGCAGAGAAAAAGATGCTCCACTATTATTAACTT 955
Db 1292 TTGATGGTCTTGTGTCATCACCATTAGCTTTCAAGAAAAACAGCTCCATTTATTATTAACTT 1351
Qy 956 CAAAAGATAAAATTAGATTTCGTCAGTAAATCTGAAATAAAGAGAGTTATGAACCTTAAAGA 1411
Db 1352 CAAAAGATAAAATTAGATTTCGTCAGTAAATCTGAAATAAAGAGAGTTATGAACCTTAAAGA 1411
Qy 1016 CTTCAACAGAGTA---ACAGGAAAAACAGTTTATATAGCTGCTGGTAGTTAATAGTAT 1072
Db 1412 GTGACACTGGTATAAATACTTCTTAAAAAGTTTATTTAGCTGGTGGAGTTAATTTCTATAT 1471
Qy 1073 CTAAGAAGATTGCTAAGCAGAAATTAGAATCAATGGGATTTAAAGTTTGAAGATTTCTCAGGTG 1132
Db 1472 CTAAGAAGTAGAANAATGAATTTGAANAACATGGGCTTTAAAGTTACTAGATTATCAGGAG 1531
Qy 1133 ATGATAGATATGAATCTTCTTTAAAAATAGCAGGTGAAATAGCTTTAGATAATATGATAAGG 1192
Db 1532 AAGACAGATACGAAACTTCTTTAGCAATAGCTGTAGAAATAGCTCTTGATATATGATAAG 1591
Qy 1193 CTTATGTAGTTGGTGAACAGGATTTAGCAGATGCCATGAGTATAGCTTTCAGTTGCTTCTA 1252
Db 1592 CATTTGTAGTTGGTGTACTGGATTTAGCAGATGCTATGAGTATAGCTCCAGTTGCTTCT- 1650
Qy 1253 CTAAATTTAGTGGTAAATGTTGTAGATAGAAACAAATGGACATGCTACTCCAAATAGTTG 1312
Db 1651 -----CAACTTAAAGATGGAGATGCTACTCCAAATAGTAG 1684
Qy 1313 TTGTAGATGGAAGAACTGATAAAATATCTGATGACTTAGATAGTTTCTTTAGAAAGCGCTG 1372
Db 1685 TTGTAGATGGAAGAAAGCAAAAGAAATAAGTGATGATGCTTAAAGAGTTTCTTAGGAACTTCTG 1744
Qy 1373 ATGATAGATATATAGTGGATTTGCAAGTGTATCTGAAAAGAGATGAAGAGCTATATCAG 1432
Db 1745 ATGTTGTATATAATAGTGGAAAAAATAGCGTATCTTAAAGAGATTGAAGAGTCAATAGATA 1804
Qy 1433 ATGCTACTGGTAAAGCGTTTACAAGATTAAAGCGCAGCATAGACAAGACACTAACTCTG 1492
Db 1805 GTGCAACTGGAAGAAACTCCAGATAGATAAGTGGAGATGATAGACAAGCAACTAATGCTG 1864
Qy 1493 AAGTTATAAAAACATATTTATGCTTAATGATCTGAAATAGCTTAAAGCTGCAGTTTATAGATA 1552
Db 1865 AAGTTTAAAGAGATGATTATTTTC----- 1890
Qy 1553 AAGATTCAGGTGCTTCAAGTAGTAGTCAGGAGTATTTAAATTTCTATGTAGCTAAAGATG 1612
Db 1891 -----ACAGATGGTGAAGTTGTGAATTTACTTTTGTGCAAAAGATG 1930
Qy 1613 GATCTACAAAAGAGATCAATTTAGTTGATGCAATTAGCAGTAGGAGCTGTTGCTGG----- 1667
Db 1931 GTTCTACTAAGAGAGATCAATTTAGTAGATGCCCTTAGCAGCAGCAACAATAGCAGGTAGAT 1990
Qy 1668 ----ATAAATACTGCTCCAGTTGTTATGCTACTGATTTCTTTATTTCTTCTGATCAATCGG 1723
Db 1991 TTAAGAGTCTCCAGCTCCAATCATACTAGCTACTGATACTTATCTTCTGACCAAAATG 2050
```

QY 1724 TTGCTATAACAAAGTTGTAGGAGAAAAATATTTCTAAAGATTTTAAACACAAAGTTGGTCAAG 1783
 Db 2051 TAGCTGTAGTAAGCAGTTCTTAAAGATGGTGGAACTAACTTAGTTCAAGTAGGTAAAG 2110
 QY 1784 GAATAGCTAATTCAGTTTATAACAAATGAAGATTTATTTAGATATG 1830
 Db 2111 GTATAGCTTCTCAGTTTATAACAAATGAAGATTTATTTAGATATG 2157

RESULT 11

US-10-068-870-5
 ; Sequence 5, Application US/10068870
 ; Publication No. US20030054009A1
 ; GENERAL INFORMATION:
 ; APPLICANT: The Provost, Fellows & Scholars of the College of the Holy and Undivided
 ; APPLICANT: Trinity of Queen Elizabeth, near Dublin
 ; TITLE OF INVENTION: C.difficile vaccine
 ; FILE REFERENCE: TRI002/C/WO
 ; CURRENT APPLICATION NUMBER: US/10/068,870
 ; CURRENT FILING DATE: 2002-02-11
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 5
 ; LENGTH: 2158
 ; TYPE: DNA
 ; ORGANISM: Clostridium difficile
 ; US-10-068-870-5

Query Match 25.7%; Score 470.2; DB 14; Length 2158;
 Best Local Similarity 61.4%; Pred. No. 2.9e-64;
 Matches 950; Conservative 0; Mismatches 468; Indels 129; Gaps 7;

QY 296 CTCAGCTAATGGAATGAAGATTTATGTAAGCAACTTTTAAATAATTTAGATGCGAGG 355
 Db 728 CAACACCTAGTGTCTGTAGCTGTAGTGGTTTTGTAACTTAAAGATGATCTGATTTAGCAA 787
 QY 356 AATATGCTATTATAGATTAACTTATATATATGCTAAACTGTGTGAAATTTAAAGTAGTAG 415
 Db 788 AATCAGGTACTTAATGTAAGAGTTTAAATGCAAAAGAGATCAATTTGATATAGATG 847
 QY 416 CAGCTAGTGAATAACAGTAGTTGTATCTAGTGTATGCAAAATATGTCGCAAAAGATATAG 475
 Db 848 CAAGCTCATATACATCAGCTG-----AAATTTAGCTAAAGATATGATTTGATCCAG 901
 QY 476 CTGAAAAATATGTTTTGAAGCAAAAGCTTAGAATATGCACTTAAACTATATAATGCTT 535
 Db 902 ATGAAATTTCTGAAGCATATTAAGCAATAGTAGCATTTACAAATATGATGTTAGAGTCTA 961
 QY 536 CAGATTTTCAGTAAACTGATAGTTACTATCAAGTAGTTCTTTATCCAAAAAGGAAAGAT 595
 Db 962 ATTTAGTTTCAGTTAGTTAATGGAATATCAAGTATTTTATCCAGAGGTAAAGAT 1021
 QY 596 TACAAGGTTTCTCAACTTATAGAGCTACAAATATTAATGAAGGAATGTCATATGGTAATA 655
 Db 1022 TAGAACTAATACAGCA-----AATGTACAAATAGCTAGTCAAGATA 1063
 QY 656 CACAGTAATATTAATCTTAAATCTACTAGTAAGATTAATTTAAAGCTGCGATAGAG 715
 Db 1064 CACAGCTAAAGTAGTTTATAAAGCTAATAATTTAAAGATTTTAAAGATTTATGTAGATG 1123
 QY 716 AGTTTCAAAATTCGATGCTAGTTTATTTCTAATCTACAACTTTTAGCTGGTGATCAGAGAA 775
 Db 1124 ATTTTAAACATATAAATAACTTATTTCAAATGTTTGAACAGTAGGAGGAGATGAAG 1183
 QY 776 TACAAACAGCTATAGAGATAAGTAAGTAATTAATAATATGATGGCGAGAAATCAGATC 835
 Db 1184 TAGAACTGCTATAGAAATTAAGTAGTAATAATTTATTTCTGAT-----GATA 1231
 QY 836 ATTCAGCTGATGTTTAAAGAGAAATGTTTAAATGTTTATTTAGTAGGTGCGAAATGCTAG 895
 Db 1232 AAAATGCAATAACTATGAAGCAGTTTATGATATAGATATTTAGTTGGATCTACATCTATAG 1291
 QY 896 TAGATGGATTAGTTGGGCTCTTTAGCAGCAGAAAAAGATGCTCCACTATTATTAATTT 955

Db 1292 TTGATGGTCTTGTGCATCACCATTTAGCTTCAGNAAAAACAGCTCCATTTATTATTAGCTT 1351
 QY 956 CAAAGATATAATTTAGATTTCTGAGTAAATCTGAAATTAAGAGAGTTTTTAGACTTAAAAA 1015
 Db 1352 CAAAGATATAATTTAGATTTCTGAGTAAATCTGAAATTAAGAGAGTTTATGAACTTAAAG 1411
 QY 1016 CTTCAACAGAGAGTA---ACAGGAAAAACAGTTTATATATAGCTGGTGGAGTTAATAGTAT 1072
 Db 1412 GTGACACTGGTATAAATACTTCTTAAAAAGATTTATTTAGCTGGTGGAGTTAATTTCTATAT 1471
 QY 1073 CTAAGAAGTTGTAAACAGAAATTTAGATCAATGGGATTTAAAGTTTGAAGATTTCTCAGGTG 1132
 Db 1472 CTAAGAAGTTGTAAAGATTAATTTGAAGAAACATGGGCTTTAAAGTTTACTAGATTTATCAGGAG 1531
 QY 1133 ATGATAGATATCAAACTCTTTTAAAAATAGCAGGTGAAATAGGCTTTAGATAATGATAAGG 1192
 Db 1532 AAGACAGATACGAACTCTTTTAGCAATAGCTGATGAATAGGCTCTTGATATGATAAG 1591
 QY 1193 CTTATGATGTTGGTGAACAGAGTTAGCAGATGCCATGATAGTATAGCTTCAGTTGCTTCTA 1252
 Db 1592 CATTTGTTAGTTGGTGGTCTGCTAGTATGATGATGATGATGATGATGATGATGATGATG 1650
 QY 1253 CTAATTTAGATGGTAAATGTTGTAGATAGAACAAATGGAATGCAATGCTACTCCAAATAGTTG 1312
 Db 1651 -----CAACTTAAAGATGGAGATGCTACTCTCCAAATAGTAG 1684
 QY 1313 TTGTAGATGGAAGCTGATATAAATATCTGATGACTTTAGATAGTTTCTTAGGAAGCGCTG 1372
 Db 1685 TTGTAGATGGAAGCAAGCAAAAGAAATTAAGTATGATGATGATGATGATGATGATGATGATG 1744
 QY 1373 ATGTAGATATAATAGTGGATTTTGCAAGTGTATCTGAAAAGATGGAAGAGCTTATATCAG 1432
 Db 1745 ATGTTGATATAATAGTGGTGAAGAAATAGCGTATCTTAAAGAGATTGAAGAGTCAATAGATA 1804
 QY 1433 ATGCTACTGGTAAAGCGGTTACAAGATTAAAGGCCACGATAGACAGACACTAACTCTG 1492
 Db 1805 GTGCAACTGGAAAACTCCAGATAGAATAAGTGGAGATGATAGACAGCAAGCACTAAATGCTG 1864
 QY 1493 AGTTTATAAAACATATTTATGCTTAATGATACTGAAATAGCTAAAGCTGCAGTTTTAGATA 1552
 Db 1865 AAGTTTTTAAAGAAAGATGATTTTTC----- 1890
 QY 1553 AAGATTTCAAGTCTTCAAGTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1612
 Db 1891 -----ACAGATGGTGAAGTGTGGAATTTACTTTGTTGCAAAAGATG 1930
 QY 1613 GATCTCAAAAGAGATCAATTTAGTTGATGATGATGATGATGATGATGATGATGATGATGATG 1667
 Db 1931 GTTCTACTAAAGAGATCAATTTAGTAGATGCTTTAGCAGCAGCAAGCAATAGCAGGTAGAT 1990
 QY 1668 -----ATATAAACTTGTCTCCAGTTGTTAGTACTGATTTCTTTATCTTCTGATCAATCGG 1723
 Db 1991 TTAAGAGTCTCCAGTCCCAATCATACTAGTACTGATTTTAAAGATTTTAAAGATTTTAAAG 2050
 QY 1724 TTGCTATAACAAAGTTGTAGGAGAAAAATATTTCTAAAGATTTTAAACAAAGTTGGTCAAG 1783
 Db 2051 TACTGTAGTAAGCAGTTTCTTAAAGATGTTGTAAGTGGTGAATCTTAGTTTCAAGTAGGTAAG 2110
 QY 1784 GAATAGCTAATTCAGTTTATAACAAATGAAGATTTTATTAGATATG 1830
 Db 2111 GTATAGCTTCTCAGTTTATAACAAATGAAGATTTTATTAGATATG 2157

RESULT 12

US-10-282-122A-16859
 ; Sequence 16859, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert


```
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16859
; LENGTH: 1893
; TYPE: DNA
; ORGANISM: Clostridium difficile
US-10-282-122A-16859

Query Match 7.9%; Score 145.2; DB 17; Length 1893;
Best Local Similarity 53.1%; Pred. No. 2.7e-13; Mismatches 21; Gaps 4;
Matches 412; Conservative 0;

QY 863 AAAATCTGTATTAGTAGGTGCAAAATGCACCTAGTAGATGGATTAGTTGGCGCTCCTTTAG 922
DB 1037 AAGCTATAATTTTGTGGAGAGATGCTATAGTTGATGGATTAGCATCAGCACCACTTG 1096
QY 923 CAGCAGAAAAGATGCTCCACTATTATTAATCTTCAAAAAGATAAATTAGATTGTCAGTAA 982
DB 1097 CCTCTCAAAAAATGCACCAATATTATTATCTAAAAAAGATTCACTACCATCAGAAATAG 1156
QY 983 AATCTGAATAAGACAGATTTTAGACTTAAAACTTCAACAGAGTAAACAGGAAAACAG 1042
DB 1157 AAGCTGAAATATTAAAGAT-----ACTTGAAGTAACTTATCTTCTTAAGAAA 1204
QY 1043 TTTATATAGCTGGTGAGTTAATAGTGTATCTAAAGAAAGTTGTAAACAGAA---TTAGAA 1099
DB 1205 TATATATAGTAGTGAGAAATCTAAAGTATCAAAAGAACTGAAGAAAACCTTCTTAAT 1264
QY 1100 CAATGGGATTAAGTTGAAAGATTTCTCAGGTGATGATAGATATGAACCTCTTTTAAAAA 1159
DB 1265 TAGGTGTAAGTAAAGTTGAGAGAGTTCTCGAGAGATAGATTGAAACTTCTTTTAGAAA 1324
QY 1160 TAGCAGGTGAATAGAGCTTAGATAAATGATAAGGCTTATAGTTGGTGAACAGGATTAG 1219
DB 1325 TAGCAAAACAAATTAAGAGATACATTTTAAGACTGCTTTTGTAGTAGTGGAAATGGAGAAG 1384
QY 1220 CAGATGCCATGATAGTCTTCAAGTTGCTTCTACTAAATTAGATGGTAAATGGTTGTTAG 1279
DB 1385 CTGATGCTATGAGTATA---TCAGCTAGAGCTGCTCAATTTGGTGTCCCAATATAGTTA 1441
```

```
QY 1280 ATAGAACAAATGGACATGCTACTCAATAGTGTGTTGTAGATGGAAGCTGATAAAATAT 1339
DB 1442 CAGGTAACGAAATTAGATGCAAAATGCTGAAAAATTTATTAAGAGGAAAGAAATTTAGA 1501
QY 1340 CTGATGACTTTAGATAGTTTCTTAGGAAGCGCTGATGTAGATATAATAGTGGATTGCAA 1399
DB 1502 TGGGTG---GAGAAAATTTCTGTATCAAAAGAAAGTTGAAGACAAATTTGTATATAG 1558
QY 1400 GTGTATCTGAAAGATGGAAGAGCTATATCAGATGCTACTGTTAAAGCGGTTACAAGAG 1459
DB 1559 TAAATAATAAGTTGAAAGATTTAGCTGGAGAAAATAGAAAAGATATACTAATGCTAAAG 1618
QY 1460 TTAAGGGCAGATAGACACACTAACTCTGAAGTTTATAAAACATATTTATGCTAATG 1519
DB 1619 TCAATATAATTAATCTATGCGAGTGCACCTTAAGCATATGTAGCAAAAGATGGTTATG 1678
QY 1520 ATACTGAAATAGCTAAAGCTGCAGTGTTTAGATAAAGATTTCAAGTGTCTTCAAGTAGT 1579
DB 1679 GAAATGGACAATTAGTTGATGCACTTACAGCAGCACCACTTGACAGCTAGTTCAAAG 1738
QY 1580 CAGGATGATTTAAATTTCTATGTAGCTAAAGATGGATCTACAAAAGAGATCAATTA 1635
DB 1739 CAATAGTATTAACTACAGAGAACTTTCTTAAATCAAGAAGAGTAGTTGAGTTA 1794
```

RESULT 13

```
US-10-282-122A-16782
; Sequence 16782, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16782
; LENGTH: 1503
; TYPE: DNA
; ORGANISM: Clostridium difficile
```



```

; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 200773USODIV
; CURRENT APPLICATION NUMBER: US/09/742,096
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 08/973,642
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: PCT/FR96/00894
; PRIOR FILING DATE: 1996-06-12
; PRIOR APPLICATION NUMBER: FR 95/07007
; PRIOR FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 5361
; TYPE: DNA
; ORGANISM: P. falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(5361)
; OTHER INFORMATION:
US-09-742-096-2

```

Query Match 5.2%; Score 95.6; DB 9; Length 5361;
 Best Local Similarity 43.6%; Pred. No. 2.2e-05;
 Matches 645; Conservative 0; Mismatches 809; Indels 24; Gaps 4;

```

QY 233 CAATGAGAAAGATTCTTCAACTCCCACTGGAGATAAAGTTTATTTCAGAACAACTT 292
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 770 CAAGTGTGAGAAAGTATAGCTTCAAGTGTGATGAAAGTATAGATTCAAGTATTGAAG 829
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 293 TAACTACAGCTAATGAAATGAAGATTATGTAAGACAA ---CTTTAAAAATTTAGATG 349
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 830 AAAATGTAGCTCAACTGTTGAAGAAATCGTAGCTCCAAGTGTGTGAAGAGTGTGGCTC 889
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 350 CAGGAGAAATGCTATATAGATTAACTTATATAATATGCTAAACTGTGTGAATTAAG 409
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 890 CAAAGTGTGAAGAAAGTGTAGAGAAATGTTGAAGAAAGTGTAGCTGAAATGTTTGAAG 949
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 410 TAGTAGCAGCTAGTGAAGAAACAGTAGTGTATCTAGTATGCGAAATATAGTGCAAAAG 469
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 950 AAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTG 1009
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 470 ATATAGCTGAAATATGTTTGAAGACAAAGACTTGAAGAAATGCACTAAAAAATATAA 529
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1010 AAAATGTTGAAGAAATCGTAGCTCCAAGTGTGGAAGAAATCGTAGCTCCAAGTGTGAAG 1069
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 530 ATGCTCAGATTTCAGTAAACTGATAGTATCTATCAAGTAGTCTCTTATCCAAAGGAA 589
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1070 AAATTGTAGCTCCAAGTGTGTGAAGAGTGTGGCTCCAAGTGTGAAGAAAGTGTAGAAG 1129
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 590 AGAGATTACAGGTTTCTCAACTTATAGAGCTACAAATATATATGAAGGAAGTGCATATG 649
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1130 AAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAG 1189
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 650 GTAATACACAGTAATATAACTCTAAAATCTACTAGTAAGAGTAAATTTAAAGACTGCGAG 709
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1190 AAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTG 1249
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 710 TAGAAGAGTTACAAAAATGCAATGCTAGTATTCTTAATCTACTCAACTTTAGCTGGTAGG 769
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1250 AAAATGTTGAAGAAATCGTAGCTCCAAGTGTGGAAGAAATCGTAGCTCCAAGTGTGAAG 1309
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 770 ACAGAAATACAAACAGCTATAGAGATAAGTAAAGAAATATTACAAATATATGATGGCGAAAT 829
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1310 AAATTGTAGCTCCAAGTGTGTGAAGAAAGTGTGGCTCCAAGTGTGTGAAGAAAGTGTAGAAG 1369
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 830 CAGATCATTCAGCTGATTTAAAGAGAAATGTTAAAAATGTTTATTTAGTAGTGCAAATG 889
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1370 AAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTTGAAG 1429
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 890 CACTAGTAGATGGAATTAGTTGGGCTCCTTTAGCAGCAGAAAAAGATGCTCCACTATTAT 949
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1430 AAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTG 1489
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 950 TAACTTCAAAAGATAAATTAGATTTCGTCAAGTAAA -----ATCTGAAATAAAGAGAG 1000
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1490 AAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAG 1549
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1001 TTTTAGACTTAAAACTTCAACAGAAAGTAAACAGGAAAAACAGTTTATATAGCTGTGTGAG 1060
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1550 AAATCGTAGCTCCAAGTGTGGAAGAAATCGTAGCTCCAAGTGTGGAAGAAATGTTAGCTC 1609
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1061 TTAATAGTGTATCTAAAGAGTGTAAACAGAAATAGAAATCAATGGGATTAAAAAGTTGAAA 1120
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1610 CAAAGTGTGTGAAGAAAGTGTGGCTCCAAGTGTGGAAGAAAGTGTAGAGAAATGTTGAAG 1669
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1121 GATTCCTCAGGTGATGATAGATATAAAGCTTCTTTAAAAATAGCAGGTGAAATAGCTTAG 1180
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1670 AAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTG 1729
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1181 ATAATGATAAGCTTTATGTGTGGGAAACAGGATTAAGCATGCCATGAGTATAGCTT 1240
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1730 AAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAATCGTAGCTCCAAGTGTGAAG 1789
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1241 CAGTTGCTTCTACTAAATTAGATGCTAATG-----GTGTTGTAGATAGAACAAATG 1291
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1790 AAATCGTAGCTCCAAGTGTGGAAGAAATGTTAGCTCCAAGTGTGTAAGAAAGTGTGGCTC 1849
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1292 GACATGCTACTCCAATAGTGTGTTAGATGGAAGAAAGCTGATAAAATATCTGATGACTTAG 1351
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1850 CAAAGTGTGGAAGAAAGTGTAGAGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAG 1909
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1352 ATAGTTTCTTAGGAAAGCGCTGATGTAGATATATATAGGTGGATTTTCAAGTGTATCTGAAA 1411
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1910 AAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAATCGTAGCTC 1969
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1412 AGATGGAAGAGCTATATACAGATGCTACTGTGTAAGGCGTTTCAAGAGTTTAAAGCGCAG 1471
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1970 CAACTGTTGAAGAAATCGTAGCTCCAAGTGTGGAAGAAATGTTAGCTCCAAGTGTGTAG 2029
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1472 ATAGCAACAGACACTAACTCTGAAGTTATAAAAAACATATTTATGCTAATGATCTGAAATAG 1531
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2030 AAAAGTGTGGCTCCAAGTGTGGAAGAAAGTGTAGAGAAATGTTTCAAGAAAGTGTG--TAG 2086
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1532 CTAAAGCTGCGAGTTTTAGATAAAGATTCAAGGTGCTTCAAGTAGTATGCCAGAGTATTTA 1591
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2087 CTGAAAAATGTTGAAGAAAGTGTAGCTGAAAAATGTTGAAGAAAGTGTAGCTGAAAAATGTTG 2146
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1592 ATTTCTATGTAGCTAAAGATGATCTACAAAGAGAGATCAATTAGTTGATGATTCAGTAGCAG 1651
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2147 AAGAAAGTGTAGCTGAAATGTTGAAGAAATCGTAGCTCCAAGTGTGGAAGAAATCGTAG 2206
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1652 TAGGAGCTGTTTCTCGATATAAAGTTGCTCCAGTTTGTGA 1689
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2207 CTCACACTGTTGAAGAAATGTTAGCTCCAAGTGTGTGA 2244
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: October 6, 2005, 11:30:56
 Job time : 3714 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM nucleic - nucleic search, using sw model
Run on: October 6, 2005, 05:21:23 ; Search time 7980 Seconds
(without alignments)
11111.910 Million call updates/sec

Title: US-10-068-870A-4
Perfect score: 1830
Sequence: 1 agaaaaaagaatttagc.....tgaaagatttagatag 1830

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues 9416466

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_btg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1830	100.0	1830	6	AX512902 Sequence
2	1815.6	99.2	1833	1	AP478570 Clostridi
3	1159.6	63.4	1182	1	AP458877 Clostridi
4	1096.2	59.9	1185	1	AP458878 Clostridi
5	546.6	29.9	2434	1	AY004256 Clostridi
6	542	29.6	2271	6	AX512904 Sequence
7	535.6	29.3	2271	1	AJ300676 Clostridi
8	535.6	29.3	2271	6	AX259458 Sequence
9	503.4	27.5	2217	6	AX512906 Sequence
10	483.2	26.4	2145	1	CDI300677 Clostridi
11	483.2	26.4	2145	6	AX259456 Sequence
12	480	26.2	2145	6	AX512907 Sequence
13	471.8	25.8	2157	6	AX512901 Sequence
14	471.8	25.8	2158	6	AX512903 Sequence
15	471.8	25.8	2158	6	AX512908 Sequence
16	471.8	25.8	2160	1	CDI291709 Clostridi
17	471.8	25.8	2160	6	AX259457 Sequence
18	470.2	25.7	2158	6	AX512905 Sequence
19	453.2	24.8	2304	1	AP478571 Clostridi

20	135	7.4	253001	3	AE014834
21	119.4	6.5	208332	10	AC122227
22	117.4	6.4	5420	3	AF056936
23	112.4	6.1	2069	6	E10125
24	112.4	6.1	3399	6	E10126
25	112.4	6.1	347050	3	PF929351
26	112.2	6.1	16377	3	AF202180
27	112	6.1	4995	3	PFAMESA
28	111.8	6.1	271546	3	AE014843
29	110.6	6.0	293650	1	AP004172
30	109.4	6.0	237549	2	AC112451
31	108	5.9	4102	3	AF270648
32	107.8	5.9	170221	10	AC127421
33	107.4	5.9	233713	2	AC133034
34	104.4	5.7	110000	2	AC129389.1
35	103.4	5.7	172307	2	AC044842
36	103.2	5.6	202920	10	AL929441
37	103	5.6	216959	10	AC116998
38	101.6	5.6	179685	10	AC122056
39	101.2	5.5	4507	1	AF093106
40	100	5.5	34347	2	BX957360
41	99.6	5.4	1833	1	AF194870
42	98.6	5.4	169614	5	BX511115
43	98.6	5.4	177122	2	BX957314
44	96.8	5.3	166529	10	AC147621
45	96	5.2	219181	2	AC141063

ALIGNMENTS

RESULT 1
AX512902
LOCUS AX512902 1830 bp DNA linear PAT 03-OCT-2002
DEFINITION Sequence 2 from Patent WO02062379.
ACCESSION AX512902
VERSION AX512902.1 GI:23504061
KEYWORDS Clostridium difficile
SOURCE Clostridium difficile
ORGANISM Clostridium difficile
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.

REFERENCE Doyle, R., Kelleher, D., Windle, H. J., Walsh, J. B. and Deirdre, N. B.
AUTHORS
TITLE Clostridium difficile vaccine
PATENT NO 02062379-A 2 15-AUG-2002;
JOURNAL THE COLLEGE OF THE HOLY AND UNDIVIDED TRINITY OF QUEEN ELIZABETH
(IE)

FEATURES Location/Qualifiers
source 1. .1830
/organism="Clostridium difficile"
/mol_type="unassigned DNA"
/db_xref="taxon:1496"

ORIGIN	Query Match 100.0%; Score 1830; DB 6; Length 1830; Best Local Similarity 100.0%; Pred. No. 4e-240; Matches 1830; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 ATCAAAAAGAAATTTAGCAATGGCTATGGCAGCTGTTACTGTAGTAGTCTGCTGCT 60
Db	1 ATCAAAAAGAAATTTAGCAATGGCTATGGCAGCTGTTACTGTAGTAGTCTGCTGCT 60
Qy	61 CCAGTTTTTGCACAGCTTCAGATGTAATATCACTACAGATGTTCAAAATGATAAGTAT 120
Db	61 CCAGTTTTTGCACAGCTTCAGATGTAATATCACTACAGATGTTCAAAATGATAAGTAT 120
Qy	121 ACAGTATCAATCACTAAAGCTAGTACTAGTAAAGGATATTTTAGCAGCACAACCTTA 180
Db	121 ACAGTATCAATCACTAAAGCTAGTACTAGTAAAGGATATTTTAGCAGCACAACCTTA 180
Qy	181 ACAACAGGTCAGTATTTTGAACAAAGATACAAAAGTTACTTTCTATGATGCAATGAG 240

Db 181 ACAACAGTGCAGTATTATTTTGAACAAAGATACAAAGTTACTTTCTATGATCAAAATGAG 240
QY 241 AAAGATTCCTCAACTCACTGAGATAAAAAAGTTTATTCAGAACTTTTAACTACA 300
Db 241 AAAGATTCCTCAACTCAACTGAGATAAAAAAGTTTATTCAGAACTTTTAACTACA 300
QY 301 GCTAATGGAATGAGATTTATGTAAGACAACTTTTAAAAAATTTAGATGTCAGAGAAATAT 360
Db 301 GCTAATGGAATGAGATTTATGTAAGACAACTTTTAAAAAATTTAGATGTCAGAGAAATAT 360
QY 361 GCTATTATAGATTTAACTTTATTAATGCTTAAACTGTTGAAATTTAAAGTAGTAGCAGCT 420
Db 361 GCTATTATAGATTTAACTTTATTAATGCTTAAACTGTTGAAATTTAAAGTAGTAGCAGCT 420
QY 421 AGTGAACAACTAGTGTCTATCTAGTGATGCGAAAAATAGTGCAAAAAGATATAGCTGAA 480
Db 421 AGTGAACAACTAGTGTCTATCTAGTGATGCGAAAAATAGTGCAAAAAGATATAGCTGAA 480
QY 481 AAATATGCTGTTGAAGACAAAGACTTTAGAAAATGCACTAAAAACTATTAATGCCTCAGAT 540
Db 481 AAATATGCTGTTGAAGACAAAGACTTTAGAAAATGCACTAAAAACTATTAATGCCTCAGAT 540
QY 541 TTCAAGTAAACCTGATAGTTACTATCAAGTAGTTCTTTTATCCAAAAAGGAAAGATTAACA 600
Db 541 TTCAAGTAAACCTGATAGTTACTATCAAGTAGTTCTTTTATCCAAAAAGGAAAGATTAACA 600
QY 601 GGTTCCTCAACTATAGAGCTCAAAATTAATTAAGAGAACTGCGATATGTTAATACACA 660
Db 601 GGTTCCTCAACTATAGAGCTCAAAATTAATTAAGAGAACTGCGATATGTTAATACACA 660
QY 661 GTAATTAATTAACCTCTAACTACTAGTAGAGCTTAATTTAAAGACTGCGATAGAGAGTTA 720
Db 661 GTAATTAATTAACCTCTAACTACTAGTAGAGCTTAATTTAAAGACTGCGATAGAGAGTTA 720
QY 721 CAAAAATGCAATGCTAGTTATTTCTAATACTACAACCTTTAGCTGGTGATGACAGAAATACA 780
Db 721 CAAAAATGCAATGCTAGTTATTTCTAATACTACAACCTTTAGCTGGTGATGACAGAAATACA 780
QY 781 ACAGCTATAGAGATTAAGTAAAGATTAATTAATATATGATGGCGAGAAATCAGATCAATCA 840
Db 781 ACAGCTATAGAGATTAAGTAAAGATTAATTAATATATGATGGCGAGAAATCAGATCAATCA 840
QY 841 GCTGATGTTAAAGAGAACTGTTAAATGTTGTTATAGTAGTGCGAAATGCACTAGTAGAT 900
Db 841 GCTGATGTTAAAGAGAACTGTTAAATGTTGTTATAGTAGTGCGAAATGCACTAGTAGAT 900
QY 901 GGATTAGTTGCGGCTCCTTTAGCAGCAGAAAAAGATGCTCCACTATTATTAACTTCAAAA 960
Db 901 GGATTAGTTGCGGCTCCTTTAGCAGCAGAAAAAGATGCTCCACTATTATTAACTTCAAAA 960
QY 961 GATAAATTAGATTCGTCAGTAAATCTGAAATAAAGAGAGTTTATAGACTTAAAAAATTTCA 1020
Db 961 GATAAATTAGATTCGTCAGTAAATCTGAAATAAAGAGAGTTTATAGACTTAAAAAATTTCA 1020
QY 1021 ACAGAGTTAAACAGGAAAAACGTTTATATAGCTGCTGGAGTTAAATAGTGTATCTAAAGAA 1080
Db 1021 ACAGAGTTAAACAGGAAAAACGTTTATATAGCTGCTGGAGTTAAATAGTGTATCTAAAGAA 1080
QY 1081 GTTGTAAACAGATTTAGATCAATGGGATTTAAAGTTTGAAGATTTCTCAGGTGATGATAGA 1140
Db 1081 GTTGTAAACAGATTTAGATCAATGGGATTTAAAGTTTGAAGATTTCTCAGGTGATGATAGA 1140
QY 1141 TATGAAACTTCTTTAAAAATAGCAGGTAAATAGGCTTAGATAAATGATAAGGCTTATGTA 1200
Db 1141 TATGAAACTTCTTTAAAAATAGCAGGTAAATAGGCTTAGATAAATGATAAGGCTTATGTA 1200
QY 1201 GTTGTGGAACAGGATTTAGCAGATGCCATGAGTATAGCTTCAAGTTGCTTCTACTAAATTA 1260
Db 1201 GTTGTGGAACAGGATTTAGCAGATGCCATGAGTATAGCTTCAAGTTGCTTCTACTAAATTA 1260
QY 1261 GATGGTAATGGTGTGTAGATAGAACAAATGGACATGCTACTCCCAATAGTTGTTGTAGAT 1320
Db 1261 GATGGTAATGGTGTGTAGATAGAACAAATGGACATGCTACTCCCAATAGTTGTTGTAGAT 1320

QY 1321 GGAAAGCTGATAAAAATATCTGATGACTTAGATAGTTTCTTAGGAAGCGCTGATGTAGAT 1380
Db 1321 GGAAAGCTGATAAAAATATCTGATGACTTAGATAGTTTCTTAGGAAGCGCTGATGTAGAT 1380
QY 1381 ATATAGCTGATTTGCAAGTGATCTCGAAAGATGGAAGAGCTATATCAGATGCTACT 1440
Db 1381 ATATAGCTGATTTGCAAGTGATCTCGAAAGATGGAAGAGCTATATCAGATGCTACT 1440
QY 1441 GGTAAAGCGCTTACAAGAGTTTAAAGCGCAGATAGACAAGACACTAACTCTGAAGTTATA 1500
Db 1441 GGTAAAGCGCTTACAAGAGTTTAAAGCGCAGATAGACAAGACACTAACTCTGAAGTTATA 1500
QY 1501 AAAACATATTAATGCTAATGATGACTGAAATAGCTAAAGCTGAGTTTATAGATAAAGATTCA 1560
Db 1501 AAAACATATTAATGCTAATGATGACTGAAATAGCTAAAGCTGAGTTTATAGATAAAGATTCA 1560
QY 1561 GGTGCTTCAAGTAGTGATGACAGGAGTATTTAATTTCTATGTAGCTAAAGATGGATCTACA 1620
Db 1561 GGTGCTTCAAGTAGTGATGACAGGAGTATTTAATTTCTATGTAGCTAAAGATGGATCTACA 1620
QY 1621 AAAGAAGATCAATTAGTTGATGCTATGATGATGATGATGATGATGATGATGATGATGAT 1680
Db 1621 AAAGAAGATCAATTAGTTGATGCTATGATGATGATGATGATGATGATGATGATGATGAT 1680
QY 1681 CCAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
Db 1681 CCAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
QY 1741 GTAGGAGAAAAATATTTTAAAGATTTTAAACAGAGTTGGTCAAGGAATAGCTAAATTCAGTT 1800
Db 1741 GTAGGAGAAAAATATTTTAAAGATTTTAAACAGAGTTGGTCAAGGAATAGCTAAATTCAGTT 1800
QY 1801 ATAAACAAAATGAAAGATTTTATTAGATG 1830
Db 1801 ATAAACAAAATGAAAGATTTTATTAGATG 1830
RESULT 2
AF478570 1833 bp DNA linear BCT 27-JUN-2002
LOCUS Clostridium difficile surface layer protein A precursor, gene,
complete cds.
ACCESSION AF478570
VERSION AF478570.1 GI:21311577
KEYWORDS
SOURCE Clostridium difficile
ORGANISM Clostridium difficile
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 1833)
AUTHORS Calabi,E. and Fairweather,N.
TITLE Patterns of sequence conservation in the S-Layer proteins and
related sequences in Clostridium difficile
J. Bacteriol. 184 (14), 3886-3897 (2002)
JOURNAL 22077258
MEDLINE 12081960
REFERENCE 2 (bases 1 to 1833)
AUTHORS Calabi,E. and Fairweather,N.F.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-2002) Biological Sciences, Imperial College,
Exhibition Rd, London, Greater London SW7 2AZ, UK
FEATURES
source
1. .1833
/organism="Clostridium difficile"
/mol_type="genomic DNA"
/strain="167"
/db_xref="taxon:1496"
/note="SlpA precursor"
/codon_start=1
/transl_table=11
/product="surface layer protein A precursor"
CDS

/protein id="AAM46789.1"
/db_xref="GI:21311578"
/translation="MKRLNLAAMAAYVTVGSAAVPAFAASDVLSIQGTNDKTVTSN
TKASDLVKDILAAQNLITGAVILINKOTKVPFDANEKUSSTFTGKKVSEYQILITAN
GNEDVYKTLKNDAGEYAIIDLTNNAKTVIYKVAASEKTVYVSSDAKNSEADIAE
KVFEDKJLENALITINADSFKTSYQVLYPKRQLQGFSTYRANKNNEGTAYGN
TPVILTKTSKSNLKTAVBEQLKLNASYNTTLAGDDRIQTALIEISKYNNNDGEK
ODHSADVKNVNVLVANALVDGLVAAPLAAEKDAPLLLTSKDLDSVSKSEIKRV
LDLKTSTVTKTVLGGVSVSKVETLESMGLKVERFSGDDRYETSLKIADEIG
LNDKAVVGGTGLADAMSIVASVSTKLDNGVVDKNGHATPIVVDGKADKISDDL
DSFLGSADVDIIGGASVSEKMEELISDLATGKGVTRKGDQDRDQNTSEVIKTYIANDT
ETAKAVALDGSSSDAGVFNFYAKDGSKEBDQLVDALAVGAVAGYKCLAPVVLAT
DSLSDQSVAISKVGEKYSKDLTVQVQGIANSINKIKDLIDM"

ORIGIN

Query Match 99.2%; Score 1815.6; DB 1; Length 1833;
Best Local Similarity 99.5%; Pred. No. 3.7e-238;
Matches 1821; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ATGAAAAAGAAATTTAGCAATGGCTATGCGAGCTGTACTGTAGTAGGTCTGCTGCT 60
DB 1 ATGAAAAAGAAATTTAGCAATGGCTATGCGAGCTGTACTGTAGTAGGTCTGCTGCT 60
QY 61 CCAAGTTTTCAGCAGCTTCAGATGTAAATATCACTCAAGATGGTACAAATGATAAGTAT 120
DB 61 CCAAGTTTTCAGCAGCTTCAGATGTAAATATCACTCAAGATGGTACAAATGATAAGTAT 120
QY 121 ACAGTATCAAAATCTAAAGCTAGTGAATAGTAAAGGATATTTTACGAGCACAAACTTGA 180
DB 121 ACAGTATCAAAATCTAAAGCTAGTGAATAGTAAAGGATATTTTACGAGCACAAACTTGA 180
QY 181 ACAACAGGTGCAGTTATTTTGAACAAAGATACAAAGTTACTTTCTATGATGCAAAATGAG 240
DB 181 ACAACAGGTGCAGTTATTTTGAACAAAGATACAAAGTTACTTTCTATGATGCAAAATGAG 240
QY 241 AAGAGTTCTTCAACTCCAGTGAATAAAGTTTATTCAGAACAACTTTTAACTACA 300
DB 241 AAGAGTTCTTCAACTCCAGTGAATAAAGTTTATTCAGAACAACTTTTAACTACA 300
QY 301 GCTAATGGAAATGAAGATATGTAAGACAACTTTTAAAGATTTAGATGCGAGAGATAT 360
DB 301 GCTAATGGAAATGAAGATATGTAAGACAACTTTTAAAGATTTAGATGCGAGAGATAT 360
QY 361 GCTATTATAGATTTAACTTAAATGCTTAAAGCTGTGCAAAATTAAGTAGTAGCAGCT 420
DB 361 GCTATTATAGATTTAACTTAAATGCTTAAAGCTGTGCAAAATTAAGTAGTAGCAGCT 420
QY 421 AGTGAAAAAAGCAGTAGTTGTATCTAGTGTGCGAAAAATAGTGCAAAAATATAGCTGAA 480
DB 421 AGTGAAAAAAGCAGTAGTTGTATCTAGTGTGCGAAAAATAGTGCAAAAATATAGCTGAA 480
QY 481 AAATATGTTTGAAGACAAAGACTTGAAGAAATGCACTAAAACTATAAATGCTCAGAT 540
DB 481 AAATATGTTTGAAGACAAAGACTTGAAGAAATGCACTAAAACTATAAATGCTCAGAT 540
QY 541 TTCAAGTAAACTGATAGTTACTATCAAGTAGTTCTTTATCCAAAAGGAAAGAGATTACAA 600
DB 541 TTCAAGTAAACTGATAGTTACTATCAAGTAGTTCTTTATCCAAAAGGAAAGAGATTACAA 600
QY 601 GGTTCCTCAACTATAGAGCTACAAATTTAATAAGAGAACTGCATATGTTAATACACCA 660
DB 601 GGTTCCTCAACTATAGAGCTACAAATTTAATAAGAGAACTGCATATGTTAATACACCA 660
QY 661 GTAATATTAACTCTAAATCTTACTAGTAAAGATTAATTAAGACTGCGAGTAGAAGAGTTA 720
DB 661 GTAATATTAACTCTAAATCTTACTAGTAAAGATTAATTAAGACTGCGAGTAGAAGAGTTA 720
QY 721 CAAAAATTTGAATGCTAGTTATCTAATACTAACACTTTAGCTGGTGTGATGACAGATACAA 780
DB 721 CAAAAATTTGAATGCTAGTTATCTAATACTAACACTTTAGCTGGTGTGATGACAGATACAA 780
QY 781 ACAGCTATAGAGATAAGTAAAGATATTTACAAATATGATGCGAGAAATCAGATCATTTCA 840

LOCUS AF458877 1182 bp DNA linear BCT 07-JUL-2002
DEFINITION Clostridium difficile strain ATCC 43594 S-layer protein variable domain SlpA (slpA) gene, partial cds.
ACCESSION AF458877
VERSION AF458877.1 GI:21702530
KEYWORDS
SOURCE
ORGANISM Clostridium difficile
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
1 (bases 1 to 1182)
Kartjalainen,T., Saumier,N., Barc,M.C., Delmee,M. and Collignon,A.
Clostridium difficile Genotyping Based on slpA Variable Region in S-Layer Gene Sequence: an Alternative to Serotyping
J. Clin. Microbiol. 40 (7), 2452-2458 (2002)
22083941
12089261
2 (bases 1 to 1182)
Kartjalainen,T.K. and Saumier,N.
Direct Submission
Submitted (16-DEC-2001) Microbiology, University of Paris-Sud,
Faculty of Pharmacy, 5, rue JB Clement, Chateauf-Malabry 92296,
France

FEATURES
Location/Qualifiers
1..1182
/organism="Clostridium difficile"
/mol_type="genomic DNA"
/strain="ATCC 43594"
/db_xref="ATCC:43594"
/db_xref="taxon:1496"
/note="serogroup: A1"
<1..>1182
/gene="slpA"
<1..>1182
/gene="slpA"
/codon_start=1
/transl_table=11
/product="S-layer protein variable domain SlpA"
/protein_id="AAW75941.1"
/db_xref="GI:21702531"
/translation="AAPFPAASDVISLQDGTNDRKYTVSNTKASDLVKDILAAQNLTT
GAVLNKDTVFYDANEKDSPTGDKVYSEOTLTANGNEDYVTKLNLDAEY
ATIDLTNNAKTVEIKVVAASEKTVVSSDKNSAKIAEKYVREDKLEKWLKTN
ASDFKTSYQVVLVYPKRKLQGFSTRATYNGTYGNTPVILTLKSTSKNLKT
AVEELQKLVNSNTITLAGDRIQTAIEISKBYNNDSKSDHSDAKVENKVVLY
GANLVGDLVLAAPLAAEKDAPLLTSKDLSDSEIKRVLDTSTEVTKTVYIA
GGVNSVAKEVVTELESMGLKVERFSGDDRYETSLKIADEIGLDNDKAYVVGGTGLADA
MS"

ORIGIN
Query Match 63.4%; Score 1159.6; DB 1; Length 1182;
Best Local Similarity 99.4%; Pred. No. 9.1e-149;
Matches 1175; Conservative 0; Mismatches 4; Indels 3; Gaps 1;
QY 55 GCTGCTCCAGTTTTCGACGAGCTTCAGATGAATATCACTACAGATGGTACAAATGAT 114
DB 1 GCTGCTCCAGTTTTCGACGAGCTTCAGATGAATATCACTACAGATGGTACAAATGAT 60
QY 115 AAGTATACAGTATCAAACTAAAGCTAGTGAATAGTAAAGGATATTTTAGCAGACAA 174
DB 61 AAGTATACAGTATCAAACTAAAGCTAGTGAATAGTAAAGGATATTTTAGCAGACAA 120
QY 175 AACTTAAACAAGGTCAGTTATTTTGAACAAAGATACAAAAGTTACTTTCTATGATGA 234
DB 121 AACTTAAACAAGGTCAGTTATTTTGAACAAAGATACAAAAGTTACTTTCTATGATGA 180
QY 235 AATGAGAAAGATCTTCAACTCACTCGAGATATAAAGTTTATTCAGAACAACTTTA 294
DB 181 AATGAGAAAGATCTTCAACTCACTCGAGATATAAAGTTTATTCAGAACAACTTTA 240
QY 295 ACTACAGCTAAATGGAATGAAGATATATGTAAGACAACTTTTAAAAAATTTAGATGCA 354
DB 241 ACTACAGCTAAATGGAATGAAGATATATGTAAGACAACTTTTAAAAAATTTAGATGCA 300

QY 355 GAATATGCTATTATAGATTAACTTATATAATATGCTAAAACTGTGAAATTAAGTAGTA 414
DB 301 GAATATGCTATTATAGATTAACTTATATAATATGCTAAAACTGTGAAATTAAGTAGTA 360
QY 415 CGAGCTAGTGAAGAAAAACAGTAGTTGTATCTAGTGTATGCGAAAAATAGTGCAGAAAGATATA 474
DB 361 CGAGCTAGTGAAGAAAAACAGTAGTTGTATCTAGTGTATGCGAAAAATAGTGCAGAAAGATATA 420
QY 475 GCTGAAAAATATGTTGTTTGAAGACAAAGACTTTAGAAAAATGCGCCACTAAAAAATATAAAT 531
DB 421 GCTGAAAAATATGTTGTTTGAAGACAAAGACTTTAGAAAAATGCGCCACTAAAAAATATAAAT 480
QY 532 GCCTCAGATTTTCAGTAAAACTGATGTTTACCTATCAAGTAGTTCTTTATCCAAAAAGGAAG 591
DB 481 GCCTCAGATTTTCAGTAAAACTGATGTTTACCTATCAAGTAGTTCTTTATCCAAAAAGGAAG 540
QY 592 AGATTACAAAGTTTCTCAAACTTATAGAGCTACAAATATATAATGAAGAACTGCATATCGT 651
DB 541 AGATTACAAAGTTTCTCAAACTTATAGAGCTACAAATATATAATGAAGAACTGCATATCGT 600
QY 652 AATACACAGTAAATATTAACCTCTAAAACTACTAGTAAAGTAAATTTTAAAGACTGCAGTA 711
DB 601 AATACACAGTAAATATTAACCTCTAAAACTACTAGTAAAGTAAATTTTAAAGACTGCAGTA 660
QY 712 GAAGAGTTACAAAAATTCGAATGCTAGTTTATTTCTAATACTACAACTTTAGCTGGTGATGAC 771
DB 661 GAAGAGTTACAAAAATTCGAATGCTAGTTTATTTCTAATACTACAACTTTAGCTGGTGATGAC 720
QY 772 AGAATACAAAACAGCTATAGAGATAAGTAAAGAAATATTACAATATATGATGGCGAGAAATCA 831
DB 721 AGAATACAAAACAGCTATAGAGATAAGTAAAGAAATATTACAATATATGATGGCGAGAAATCA 780
QY 832 GATCAATTCAGCTGATGTTTAAAGAGAAATGTTTAAAAATGTTTATAGTAGTGGTGCAAAATGCA 891
DB 781 GATCAATTCAGCTGATGTTTAAAGAGAAATGTTTAAAAATGTTTATAGTAGTGGTGCAAAATGCA 840
QY 892 CTAGTAGTAGTATGTTGCGCTCTTTAGCAGCAGAAAAAGATGCTCCACTATTATTA 951
DB 841 CTAGTAGTAGTATGTTGCGCTCTTTAGCAGCAGAAAAAGATGCTCCACTATTATTA 900
QY 952 ACTTCAAAAGATAAATAGATTTCGTCACTAAAAATCTGAAATATAAGAGAGTTTATAGACTTTA 1011
DB 901 ACTTCAAAAGATAAATAGATTTCGTCACTAAAAATCTGAAATATAAGAGAGTTTATAGACTTTA 960
QY 1012 AAAAATCTCAACAGAGTAAACAGGAAAAACAGTTTATATAGCTGGTGGAGTTAATAGTGA 1071
DB 961 AAAAATCTCAACAGAGTAAACAGGAAAAACAGTTTATATAGCTGGTGGAGTTAATAGTGA 1020
QY 1072 TCTAAGAGAGTTGTAAACAGAAATTAAGATCAATGGATTTAAAGTTTGAAGATTTCTCAGGT 1131
DB 1021 GCTAAGAGAGTTGTAAACAGAAATTAAGATCAATGGATTTAAAGTTTGAAGATTTCTCAGGT 1080
QY 1132 GATGATAGATGATAAACTTTCTTTAAAAATAGCAGGTGAAATAGGCTTAGATAAATGATAAG 1191
DB 1081 GATGATAGATGATAAACTTTCTTTAAAAATAGCAGATGAATAGGCTTAGATAAATGATAAG 1140
QY 1192 GCTTATGATGTTGGTGGAAACAGGATTAGCAGATGCCATGAGT 1233
DB 1141 GCTTATGATGTTGGTGGAAACAGGATTAGCAGATGCCATGAGT 1182
RESULT 4
AF458878
LOCUS
DEFINITION Clostridium difficile strain TO005 S-layer protein variable domain SlpA (slpA) gene, partial cds.
ACCESSION AF458878
VERSION AF458878.1 GI:21702532
KEYWORDS
SOURCE
ORGANISM Clostridium difficile
Clostridium difficile
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

Clostridium.
1 (bases 1 to 1185)
Karjalainen, T.; Saumier, N., Barc, M.C., Delmee, M. and Collignon, A.
Clostridium difficile Genotyping Based on *slpA* Variable Region in
S-Layer Gene Sequence: an Alternative to Serotyping
J. Clin. Microbiol. 40 (7), 2452-2458 (2002)
22083941
MEDLINE
PUBMED
2 (bases 1 to 1185)
Karjalainen, T.K. and Saumier, N.
Direct Submission
Submitted (16-DEC-2001) Microbiology, University of Paris-Sud,
Faculty of Pharmacy, 5, rue JB Clement, Chatenay-Malabry 92296,
France
FEATURES
source
Location/Qualifiers
1. .1185
/organism="Clostridium difficile"
/mol_type="genomic DNA"
/strain="TO005"
/db_xref="taxon:1496"
/note="serogroup: A10"
<1. .>1185
/gene="slpA"
<1. .>1185
/gene="slpA"
/codon_start=1
/transl_table=11
/product="S-layer protein variable domain *SlpA*"
/protein_id="AA075942.1"
/db_xref="GI:21702533"
/translations="AAPVFAAASDVLSQDTNDKVTNSKASDLVKDILAQNLT
GALCRQDTSYFLCQMKWSSTATLEIKEYVSEIOLLTANGNEDYVKTKLNLGAD
EYALIDLTNNAKTEIKVVAASEKTVVSSDAKNSAKDIAEKYVPEKDLLENALAKI
NASDFSKTSDIYQVLLPKGRQGSFTKATNNEGTAICVTPVILTLKSKSKNLK
TAYVEIQKLNASYNTTLGDDRIQTAIBISKEYNNQGEKDHSDKVENKVNVL
VGNALVDGIVAAPLAAEKDAPLLTSDKLDSSVKSEIKRVLDTLKTSTEVTKTVYI
AGGVNSVSKVWTELESMLKVERFSGDDRYETSLKIADIEGLDNKAYVVGTTGLAD
AMS"

ORIGIN
Query Match 59.9%; Score 1096.2; DB 1; Length 1185;
Best Local Similarity 97.6%; Pred. No. 3.9e-140;
Matches 1156; Conservative 0; Mismatches 23; Indels 6; Gaps 4;
55 GCTGCTCAGTTTTCGACGAGCTTCAGATGTAATATCCTACAGAGTGTCAAAATGAT 114
Db 1 GCTGCTCAGTTTTCGACGAGCTTCAGATGTAATATCCTACAGAGTGTCAAAATGAT 60
115 AAGTATACAGTATCAAAATCTAAAGCTAGTGTAGTAAAGGATATTTTAGCAGACAA 174
Db 61 AAGTATACAGTATCAAAATCTAAAGCTAGTGTAGTAAAGGATATTTTAGCAGACAA 120
175 AACTTAACAACAGTGTGATTTTGAACAAGATACA-AAAGTTACTTTCTATG-ATG 232
Db 121 AACTTAACAACAGTGTGATTTGTAACAAGGATACAGAAAGTACTTTCTATGATG 180
233 CAATGAGAAAGAT-TCCTGACCTCAACTGCGA- -TAAAGAGTTTATTCAGACAA 288
Db 181 CAATGAGAAAGATGCTTCAACTGCCACACTGGAGATAAAGAGTTTATTCAGACAA 240
289 ACTTTACTACAGCTAATGGAATGAGATATGTAAGACAACTTTAAAGAAATTTAGAT 348
Db 241 ATTTTAACTACAGCTAATGGAATGAGATATGTAAGACAACTTTAAAGAAATTTAGAT 300
349 GCAGGAGATATGCTATTATAGATTTTAACTTATTAATATGCTAAAGCTTTGAAATTTAA 408
Db 301 GCAGGAGATATGCTATTATAGATTTTAACTTATTAATATGCTAAAGCTTTGAAATTTAA 360
409 GTAGTAGCAGCTAGTGAAGAAACAGTAGTGTGATGATGCGAAATAGTGCAAAA 468
Db 361 GTAGTAGCAGCTAGTGAAGAAACAGTAGTGTGATGATGCGAAATAGTGCAAAA 420
469 GATATAGCTGAAAATATGCTGTTTGAAGACAAAGACTTTAGAAATGCACATAAAACTATA 528

Db 421 GATATAGCTGAAAATATGCTGTTTGAAGACAAAGACTTAGAAAATGCCTAAAACTATA 480
Qy 529 AATGCTCAGATTTTCAGTAAACTGATAGTACTCAAGTAGTCTTTTATCCAAAGGA 598
Db 481 AATGCTCAGATTTTCAGTAAACTGATAGTACTCAAGTAGTCTTTTATCCAAAGGA 540
Qy 589 AAGAGATTACAAGGTTTCTCAACTATATAGAGCTACAAATATTAATGAAGAACTGCATAT 648
Db 541 AAGAGATTACAAGGTTTCTCAACTATATAGAGCTACAAATATTAATGAAGAACTGCATAT 600
Qy 649 GGTATACACCACTAATATTAATCTTAATCTTAACTAGTAAAGTAATTTAAAGACTGCA 708
Db 601 GGTATACACCACTAATATTAATCTTAACTAGTAAAGTAATTTAAAGACTGCA 660
Qy 709 GTAGAAGAGTTACAAAATTTGAATGCTAGTATTTCTAATCTACAACTTTAGCTGGTAT 768
Db 661 GTAGAAGAGTTACAAAATTTGAATGCTAGTATTTCTAATCTACAACTTTAGCTGGTAT 720
Qy 769 GACAGATAACAAACAGCTATAGAGATAAGTAAAGATAATTAACATAATGATGCGAGAAA 828
Db 721 GACAGATAACAAACAGCTATAGAGATAAGTAAAGATAATTAACATAATGATGCGAGAAA 780
Qy 829 TCAGATCATTCAAGTGTGTTAAAGAGATGTTTAAAGATGTTCTATTTAGTGTGCAAT 888
Db 781 CAAGATCATTCAAGTGTGTTAAAGAGATGTTTAAAGATGTTCTATTTAGTGTGCAAT 840
Qy 889 GCACTAGTAGTGGATTTGCGGCTCTTTAGCAGCAGAAAAGATGCTCCACTATTA 948
Db 841 GCACTAGTAGTGGATTTGCGGCTCTTTAGCAGCAGAAAAGATGCTCCACTATTA 900
Qy 949 TTAACCTTCAAAAGATATAATTTAGATTCGTAGTAAATCTGAATAAAGAGAGTTTAGAC 1008
Db 901 TTAACCTTCAAAAGATATAATTTAGATTCGTAGTAAATCTGAATAAAGAGAGTTTAGAC 960
Qy 1009 TTAACCTTCAAAAGATATAATTTAGATTCGTAGTAAATCTGAATAAAGAGAGTTTAGAC 1068
Db 961 TTAACCTTCAAAAGATATAATTTAGATTCGTAGTAAATCTGAATAAAGAGAGTTTAGAC 1020
Qy 1069 GTATCTAAAGAGTGTGTAACAGAAATTTAGATCAATGGGATTTAAAGTTGAAGATTTCTCA 1128
Db 1021 GTATCTAAAGAGTGTGTAACAGAAATTTAGATCAATGGGATTTAAAGTTGAAGATTTCTCA 1080
Qy 1129 GGTGATGATAGATATGAAACTCTTTTAAATAAGCAGGTGAAATAGCTTTAGATAATGAT 1188
Db 1081 GGTGATGATAGATATGAAACTCTTTTAAATAAGCAGGTGAAATAGCTTTAGATAATGAT 1140
Qy 1189 AAGGCTTATGATGTTGGTGGACAGATTTAGCAGATGCCATGAT 1233
Db 1141 AAGGCTTATGATGTTGGTGGACAGATTTAGCAGATGCCATGAT 1185

RESULT 5
AY004256 2434 bp DNA linear BCT 09-APR-2001
LOCUS Clostridium difficile *slpA* (slpA) gene, complete cds.
DEFINITION
ACCESSION AY004256
VERSION AY004256.1 GI:11496149
KEYWORDS
SOURCE
ORGANISM
Clostridium difficile
Clostridium difficile
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE
1 (bases 1 to 2434)
AUTHORS Karjalainen, T., Waligora-Dupriet, A.J., Cerquetti, M., Spigaglia, P.,
Maggioli, A., Mauri, P. and Mastrantonio, P.
TITLE Molecular and genomic analysis of genes encoding surface-anchored
proteins from Clostridium difficile
JOURNAL Infect.-Immun.-69 (5)-3442-3446-(2001)
MEDLINE 1189268
PUBMED 11292772
REFERENCE 2 (bases 1 to 2434)
AUTHORS Karjalainen, T.K.

← noted Dec. 2006
R

TITLE Direct Submission
JOURNAL Submitted (05-JUL-2000) Microbiology, Univ Paris-Sud, Faculty of Pharmacy, 5, rue JB Clement, Chatenay-Malabry 92296, France
FEATURES Location/Qualifiers

source 1..2434
/organism="Clostridium difficile"
/mol_type="genomic DNA"
/strain="79685"
/db_xref="taxon:1496"
56..2209
/gene="slpA"
56..2209
/gene="slpA"
/note="S layer protein"
/codon_start=1
/transl_table=11
/product="SlpA"
/protein_id="AAF89093.1"
/db_xref="GI:11496150"
/translation="MSGLVFLASAAVPFAADVKAEEYITVQDKYKDTLKKIQAGIKDGS
ITNLVVTYDKKEVANYNYSKDSATTTADAKEIAATLNLVDSKLNLGDDLVSPNIK
YDAEKFHTKDEMDALTKLENKEIVKPASETTAGLVMADGATDSKADSKSYAKDVI
KFDVSDTTGYKLITATPIADAQLATLKATYKYANNNTKVFASATELAAIDGSAVEYAK
GKEYNATGSLVFSATGKTSNINVDPLTNKGDTVVKVINAKESTIDIDSSTGSAEDL
AKKYVFDEKLDLDIYKELTSEEGYNLVOLVSRGYOVALYPEGKRLDTKGATDIENTP
VKLVLRADKIKDKMDYIDDLRTYNNYSNVVTVAGEDRLETAIEISNKYNSDDXKAI
TSATDSVVLVGSQALVDGLVSLPASEKHPALLLTSKDKLSDNSVNSEIKRVDLKST
SGINTSKKVLVAGVNSISKEVENELDKMGLKVTLSGDDRYETSLIADEVLGDNOK
AFVGGTGLADAMSAPVASQLKSNGLDLDVVDGDAPIVVVDGKAKTINNETEDFLN
NAQVDIIGGNSVSKDVKESIIVVATGKSPNRTSGDDROATPNAEVKMETDYFPEKGSVIN
YFVAKDGSSTKEDQLDALAAAPVAANFSGTIDGKNANGTVSPAPIVLATDLSLSAQNV
GYSKSVSDDGGKRLVQVGKLIASSVSKMKDLDM"

ORIGIN

Query Match 29.9%; Score 546.6; DB 1; Length 2434;
Best Local Similarity 59.0%; Pred. No. 2.3e-65;
Matches 1061; Conservative 0; Mismatches 684; Indels 54; Gaps 5;
35 CTGTTACTGTAGTGGTTCTGCTGCTCCAGTCTTTTCGACGAGCTTCAGATGTAATATCAC 94
459 CAGCTGCTGTAGTAATGGCTGTGGTCTACTGATTTCAAAAAGCAGATATAAAGTTTAT 518
95 TACAAGATGTCACAAATGATAAGTATACAGTATCAAACTACTAAAGCTAGTGACTTAGTAA 154
519 ATGCTAAGATGTAATTAAGTTTGATGTAGTATACAGTATATAGATATAGATTTACAG 578
155 AGGATATTTTTCAGCAGCACAAACTTAACCAAGCTGCAGTATTTTGAACAAAGATACAA 214
579 CAACACCAATTGCTGTATGACAAATAGCAACTTTAAAGCTACTTTATAAATATGCAATA 638
215 AAGTTACTTCTATGATGCAAAATGAGAAAGATTCTTCAACTCCAACTGGAGATATAAAG 274
639 ATACAAAGGTAGAAATTTGCTAGTGTCTACAGAGCTTGTCTGTACAGATGGTTCAGCTGTAG 698
275 TTTTATTTCAGAACAACTTTAACTACAGCTAATCGAATGAAGATTATGTAAGACAACTT 334
699 AAGTTGCAAAAGGAAAGAAATATAATCGGACTGGCTCACTTGTATTTGATGTAGTGTACTG 758
335 TAAAAAATTTAGATGCAGGAGAAATATGCTATTATAGATTTAACTTATAATATGCTTAAAA 394
759 GAAAGACTTCTAATATAAATGTTGATCCTTCTTAAACCAAGGCTGATACAGTTGTAAAAAG 818
395 CTGTTGMAATTAAGTAGTAGAGCTAGTGAAAAACACAGTAGTTGTATCTAGTGTAGCGGA 454
819 TTATAAATGCAAAAGAAATCAACAATAGATATAGATTTCAAGTACTAGTACAAAGTGTGAAG 878
455 AAAATAGTCAAAAGATATAGCTGAAAAATATGTTGTTGAAGCAAAAGACTTAGAAAAATG 514
879 ATTTAGCTAAGAAATATGTTATTTGATGAAGATATAAATAGATATATATAAGATTAA 938
515 CACTAAAAACTATAAATGCCTCAGATTTTCAGTAAAACTGTAGTTACTATCAAGTAGTTTC 574
939 CTAGTGAAGAGGATATGAAAAATTTAGTACAACTAGTAGTGAAGTGTAGATATCAAGTAGCTC 998

575 TTTATCCAAAGGAAGAGATTACAAAGGTTTCTCAACTTATAGAGCTACAAATTTAATG 634
999 TTTATCCAAAGGAAGAGATT-----AGATACTAAAG 1031
635 AAGGAACGTGCATATCGTAAATACACCAGTAATATTAACCTCTAAAATCTACTAGTAAGATA 694
1032 GAGCAACAGATATAGAAATACTCTCTGTTAAATTAGTTCTTAAAGCAGATAGATAAAG 1091
695 ATTTAAAGACTGCAGTACAGAGATTACAAAATTTGAATGCTAGTTATTCTTAATCTACAA 754
1092 ATATCAAAAGATTATATAGATGATTATTAAGAACATATAACAACAGTATTTCAAATGTTGTA 1151
755 CTTTAGCTGGTGCATCAGACAGATACAAAACAGCTATAGAGTAAAGATATTAACAATA 814
1152 CTGTAGCTGGAGAAGATAGATAGAAACAGCTATAGAATTAAGTAAATTAATTAATTAAT 1211
815 ATGATGGCGGAGAAATCAGATCAATTCAGCTGATGTTTAAAGAGAAATGTTAAAAATGTTGAT 874
1212 CTGAT-----GATAGCATGCTATTAACCTGATTCAGCAACTGATTCAGTAGTTT 1259
875 TAGTAGGTGCAAAATGCATAGTAGATAGATTAGTTGCGGCTCCTTTAGCAGCAGAAAAAG 934
1260 TAGTTGGTCTCAAGCTATAGTTGATGGTCTTGTGTCATCACCTTTAGCATCAGAAAAAC 1319
935 ATGCTCCACTATTATTAACCTTCAAAAGATAAATTAAGATTGCTCAGTAAATCTGAAATA 994
1320 ATGCTCCCAATTATTAACCTTCAAAAGATAAATAGATTCAATGTTAAATCTGAGATAA 1379
995 ACAGAGTGTTTAGACTTTAAAACTTCAACAGAGATA---ACAGGAAAAACAGTTTATATAG 1051
1380 AAAGAGTTATGATTTTAAATCTACAAGTGGTATAAATCTTCTTAAAAAGTTTATTTAG 1439
1052 CTGTTGGAGTTAAATAGTGTATCTAAAGAAAGTTGTAACAGATTAAGATCAATCGGATTA 1111
1440 CTGTTGGAGTTAACTCTATATCTAAAGAAAGTTGAAATCAATTAAGAGATATGGACTTA 1499
1112 AAGTTGAAGAGTCTCAGGTGATGATAGTAACTTCTTAAAAATAGCAGCTGAAA 1171
1500 AAGTTACAAGATTATCTGGTGATGACAGATATGAAACTTCTTTGCTATAGCTGATGAAG 1559
1172 TAGGCTTAGATAAATGATAAGGCTTATAGTATGGTGGAAACAGGATTAGCAGATGCCATGA 1231
1560 TAGTCTTGTATGATGATAAAGCAATTTGATTTGGTGGAACTGGATTAGCAGATGCTATGA 1619
1232 GTATAGCTTCAGTTGCTTCTACTAAATTAAGTGGTAAATGTTGTTGTTAGATAGAACAAATG 1291
1620 GTATAGCTCCAGTTGCTTCTCAATTAAGAAATCTAAATGGAGATTTAGATGTAGTTGATG 1679
1292 GCATGCTACTCCAAATAGTTGTTGATAGTGGAAAGCTGATAAATAATCTGATGACTTAG 1351
1680 GAGATGCTACTCCAAATAGTTGTTGATAGTGGAAAGCTGATAAATAATTAATGAACAG 1739
1352 ATAGTTTCTTAGGAAGCGCTGATGTAGATATAATAGTGGGATTTGCAAGTGTATCTGAAA 1411
1740 AAGATTTCTTAAACAAATGCACAAGTTGATATAATAGTGGGAAACACAGTGTATCTAAG 1799
1412 AGATGGAAAGAGCTATATCAGATGCTACTGGTAAAGGGCGTTTCAAGAGTTAAAGCGGAG 1471
1800 ATGTAGAAAAATCTATAGTTGTTGCTACTGGAAAAAGAGCCTAACAGAACTAGTGGAGATG 1859
1472 ATAGACACAGACTAACTCTGAAGTTTAAACAACTATTTATGCTAACTGATCTGAAAAATG 1531
1860 ATAGACAAAGCTACTAATGAGAAATTAAGAAACCTGATTAATCTTCAAAAGGAAGATG 1919
1532 CTAAAGCTGCAGTTTTAGATAAAGATTTCAGGTGCTTCAAGTAGTAGTATGACGAGGATTTTA 1591
1920 TAAATAACTTACTTCTGAGCAAAAGAT---GGTTCTACTTAAAGAGATCAATTTAGTAGATG 1976
1592 ATTTCTATGTAGTAGTAAAGATGATCTACAAAGAGAGATCAATTAGTTGATGCAATTAGCAG 1651
1977 CTTTAGCAGCAGCTCCAGTTGAGCTAACTTCGGTTCTACTTATGATGTTAAAAACGCTA 2036

QY 1652 TAGGAGCTGTGCTGGATATAAACTTGTCCAGTGTGTTAGTACTGTATCTTTATCTT 1711
Db 2037 ATGGAACTGTTCT-----CCAGCTCCATAGTATTAGTACTGTATCTTATCTG 2087
QY 1712 CTGATCAATCGGTGCTATAAGCAAAAGTTGTAGGAGAAAATATCTTAAAGATTAAAC 1771
Db 2088 CAGACCAAAATGTAGGTGAAGTAATCACTAGTATGATGTGTGGAAGAACTTAGTTTC 2147
QY 1772 AAGTTGGTCAAGAAATAGCTAATTCAGTTATAACAAATGAAGATTATTTATAGATATG 1830
Db 2148 AAGTTGGTAAAGGTATGCTAGTTTCAAGTATAGCAAAATGAAGATTATTTATAGATATG 2206

RESULT 6
AX512904 2271 bp DNA linear PAT 03-OCT-2002
LOCUS
DEFINITION Sequence 4 from Patent WO02062379.
ACCESSION AX512904
VERSION AX512904.1 GI:23504063

KEYWORDS
SOURCE Clostridium difficile
ORGANISM Clostridium difficile
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.

REFERENCE
AUTHORS Doyle, R., Kelleher, D., Windle, H. J., Walsh, J. B. and Deirdre, N. B.
TITLE Clostridium difficile vaccine
JOURNAL Patent: WO-02062379-A 4 15-AUG-2002;
THE COLLEGE OF THE HOLY AND UNDIVIDED TRINITY OF QUEEN ELIZABETH
(IE)

FEATURES
source Location/Qualifiers
1..2271
/organism="Clostridium difficile"
/mol_type="unassigned DNA"
/db_xref="taxon:1496"

ORIGIN

Query Match 29.6%; Score 542; DB 6; Length 2271;
Best Local Similarity 59.4%; Pred. No. 9.8e-65;
Matches 1117; Conservative 0; Mismatches 670; Indels 93; Gaps 8;

QY 32 CAGCTGTTACTGTAGTAGTCTGCTGCTCCAGTCTTTTGCAGCAGCTTCAGATGTAATAT 91
Db 401 CAGCAGATGCTATTAATTTGCTGGACATCTTCAGCAGATGTTGTTTACAACTCGAG 460
QY 92 CACTACAAGATGGTACAAATGATAGTATACAGTATCAATACATAAGCTAGTACTTAG 151
Db 461 CTGCTAGTGGTCTTACTGAGACAAATTCAGCAGGACAAACTTGCATATGTCAGCTATT 520
QY 152 TAAAGGATATTTTAGCAGCACAAACTTAAACACAGGTGCGATTTATTTTGAACAAAGATA 211
Db 521 TTGACACAGCATATACAGATTTCATCTGAAACTGCGGTTAAGATTACTATATAAAGCAGATA 580
QY 212 CAAAGTTACT--TTCTATGATCAATGAGAAGATCTTCACACTCCACTCGAGATA 268
Db 581 TGAATGATACATAAATTTGGTAAAGCAGGTGAGACACTTATCACTGGGCTTACATTG 640
QY 269 AAAAAAGTTTATTCAGAACAACTTTAACTACAGCTAATGGAATGAAATGAAATGATCTAAAGA 328
Db 641 AAGATGGCTTACAGAAAATTTGTAATTTAGGGGACAGTGATATTATAGATATACTA 700
QY 329 CAACCTTAAAA--AATTAGATGAGGAGAAATATGCTATTATATAGATTAACTTATATA 385
Db 701 AAGCTCTTAAACTTACTGTTCTGCTGGAAGTAAAGCAACTGTGTTAAGTTTGTCTGAAAAA 760
QY 386 ATGCTAAACCTGTTG-----AAATTAAGTAGTAGCAGCTAGTGA 427
Db 761 CACCAAGTGCAGGTGTTCAACAGATATAACAAAGCTTAGAATTAATATGCTTAAAGAG 820
QY 428 AAACAGTAGTTGTATCTAGTGATGCGAAAAATAGTGCAAAAAGATATAGCTGAAAAATATG 487
Db 821 AAACATAGATATTGACCTAGTTCTAGTAAACAGCACAGATTTAGCTAAAAAATATG 880

QY 488 TGTTTTGAACACAAAGACTTTAGAAAATGCACTAAAACTATAAAATGCCTCAGATTTTCACTA 547
Db 881 TATTTAAATAAACTGATTAAATCTCTTTATAAAAGTATTTAAATGGAGATGAAGCAGATA 940
QY 548 AAACGTAGATTACTATCAAGTAGTCTTTTATCCAAAAGAAAGAGATTACAAAGTTTCT 607
Db 941 CTAATGGATTAAATAGAGAAGTTAGTGGAAAATATCAAGTAGTTCTTTATCCAGAAGAA 1000
QY 608 CAACTTATAGAGCTACAAATTAT--AATGAAGGAACCTGCATATGTTATATACACCACTAA 664
Db 1001 AAGAGTTACACTAAGAGTGTGCAAGGCTTCAATTCGTGATGAAAATTCACCACTTA 1060
QY 665 TATTAACCTCTAAATCTACTAGTAAGAGTAATTTAAAGACTGCGAGTAGAAGATTACAA 724
Db 1061 AATTAACCTCTAAGTCAGATAAGAAAGAAAGACTTTAAAGATTATGTGGAGTATTTAAGAA 1120
QY 725 AATTGATCTAGTTATTTCTAATCTACTACAACTTTAGCTGTGATGACAGATAACAAACAG 784
Db 1121 CATATAAATGGATATTCAAATGCTATAGAGTAGCAGAGAGAAATAGATAGAACTG 1180
QY 785 CTATAGAGATAAGTAAAGAAATATTAACAATGATGCGAGAAATCAGATCATTCAGCTG 844
Db 1181 CAATAGCATTAAGTCAAAATATTTATACTCTGATGATGAAA-----TGCTA 1228
QY 845 ATGTTAAAGAGAAATGTTAAAAATGTTGTTATTTAGTGGTCAAAATGCACTAGTAGATGGAT 904
Db 1229 TATTTAGAGATTTCAGTTGATTAATGTTAGTTGTTGGAGAAATGCAATAGTTGATGAG 1288
QY 905 TAGTTGGGCTCTTTAGCAGCAGAAAAGATGCTCCACTATTTATTAACCTTCAAAAGATA 964
Db 1289 TTGTAGCTTCTCCTTTAGCTTCTGAAAAGAAAGCTCTCTTTATTTATTAACCTTCAAAAGATA 1348
QY 965 AATTAGATTCTGTCAGTAAATCTGAAATAAGAGAGTTTTTAGACTTAAAAACTTCAACAG 1024
Db 1349 AATTAGATTCAACGGTAAAGCTGAAATTAAGAGAGTTATGATATATAAGAGTACACAG 1408
QY 1025 AAGTA---ACAGGAAAAACAGTTTTATATAGCTGGTGGAGTTAATAGTGTATCTAAAGAG 1081
Db 1409 GTATAAATACTTCAAGAAAGTTTATTTAGCTGGTGGAGTTAATTTCTATATCTTAAAGAG 1468
QY 1082 TTGTAACAGAAATTAGAAATCAATGGGATTAAGAGTTGAAAGATTCCTCAGGTGATGATAGAT 1141
Db 1469 TAGAAAAATGAATTTAAAGATATGCGACTTAAAGTTTACAAGATTAGCAGGAGATGATAGAT 1528
QY 1142 ATCAAACTTCTTTAAAAATAGCAGGTGAAATAGCTTAGATAATGATAAGGCTTATGTAG 1201
Db 1529 ATGAACTTCTCTAAAAATAGCTGATGAAGTAGTCTTGTATATGATAAGCATTTGTAG 1588
QY 1202 TTGTTGGAAACAGGATTAGCAGATGCCATGATAGTATAGCTTCAGTTGCTTCTACTAAATTAG 1261
Db 1589 TTGGAGGAACAGGNTTAGCAGATGCCATGATAGTATAGCTCCAGTTGCTATCTCAATTAAGAA 1648
QY 1262 ATGTTAATGTTGTTAGATAGAAACAAATGGACATGCTACTCCNATAGTTGTTGTAGATG 1321
Db 1649 ATGCTAATGTAATAATGGATTATAGCTGATGCTGATGCTACCAATAGTAGTTGTAGATG 1708
QY 1322 GAAAGCTGATAAATCTGATGACTTAGATAGTTTCTTAGGAAGCGCTGATCTAGATA 1381
Db 1709 GAAAGCTAATACTATATAATGATGATAAAGATTTCTTAGATGATTACAGTTGATA 1768
QY 1382 TAAATAGTGGATTTGCAAGTGTATCTGAAAAGATGGAAGAGCTTATATCAGATGCTACTG 1441
Db 1769 TAAATAGTGGAGAAACAGTGTATCTAAAGATGTTGAAAATGCAATAGATGATGATGATG 1828
QY 1442 GTAAAGCGTTACAAAGATTAAAGCGGACAGATAGACAAAGACATTAACCTCTGAAGTTATA 1501
Db 1829 GTAAATCTCCAGATAGATATAGTGGAGATGATAGACAAAGCACTAATTCGCAAAAGTTATA 1888
QY 1502 AAACATATTATGCTAATGATGACTGAAATAGCTTAAAGCTGCGAGTTTTTATAGATAAGATTCAG 1561
Db 1889 AAGAACTTCTTATTTATCAAGATAACTTAAATAATGATAAAAAAGTAGTTTATTTCTTTG 1948
QY 1562 GTGCTTCAAGTAGTAGTGACAGAGTATTTTAAATTTCTATGTAGCTTAAAGATGGATCTACAA 1621

Db 1949 TAGCTAAAGATGTTCTTACTAAAGAAGATCAATTAGTGTCTTTAGCAGCAGCTCCAG 2008
QY 1622 A-----AGAAAGATCAATTAGTTG 1639
Db 2009 TTGCAGCAACITTTGGTGTAACTCTTAATTCTGATGGTAAGCCAGTAGATAGATAAGATGGTA 2068
QY 1640 ATGCATTAGCAGTAGAGAGCTGTTGCTGGATATAAA-----CTTGCTCCAGTTGTTAT 1690
Db 2069 AAGTATTAACTGGTCTGATATAGATAAAAAATAAATTAGTATCTCCAGCACCTATAGTAT 2128
QY 1691 TAGCTACTGATCTTTTATCTTCTGATCAATCGTGTCTATAGCAAAAGTTGTAGGAGAA 1750
Db 2129 TAGCTACTGATCTTTTATCTTCTGATCAAAAGTGTATCTATAAGTAAAGTTCTTGTATAAG 2188
QY 1751 AATATTCTAAAGATTTAACACAAAGTTGGTCAAGGAATAGCTAATTCAAGTTATAAAACAAA 1810
Db 2189 ATAATCGAGAAACTTAGTTTCAAGTTGGTAAGGTATAGCTACTACTTCAGTTATAACAAA 2248
QY 1811 TGAAGATTATTAGATATG 1830
Db 2249 TGAAGATTATTAGATATG 2268

RESULT 7
CDI300676
LOCUS
DEFINITION 2271 bp DNA linear BCT 04-SEP-2001
ACCESSION
AJ300676 Clostridium difficile sfpA gene for S-layer protein, strain R8366.
VERSION
AJ300676.1 GI:13539181
KEYWORDS
S-layer protein; sfpA gene.
SOURCE
Clostridium difficile
ORGANISM
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE
1 Calabi, B., Ward, S., Wren, B., Paxton, T., Panico, M., Morris, H.,
Dell, A., Dougan, G. and Fairweather, N.
TITLE
Molecular characterization of the surface layer proteins from
Clostridium difficile
JOURNAL
Mol. Microbiol. 40 (5), 1187-1199 (2001)
MEDLINE
21295357
PubMed
11401722
REFERENCE
2 (bases 1 to 2271)
AUTHORS
Fairweather, N.F.
TITLE
Direct Submission
JOURNAL
Submitted (19-DEC-2000) Fairweather N.F., Biochemistry, Imperial
College, Exhibition Road, London, SW7 2AY, UNITED KINGDOM
FEATURES
source
1. .2271
/organism="Clostridium difficile"
/mol_type="genomic DNA"
/strain="R8366 (ribotype #1)"
/db_xref="taxon:1496"
/country="United Kingdom"
1. .2271
/gene="sfpA"
1. .2271
/gene="sfpA"
/function="surface layer protein"
/codon_start=1
/transl_table=11
/product="S-layer protein"
/protein_id="CAC35720.1"
/db_xref="GI:13539182"
/db_xref="UniProt/TREMBL:Q9AEM3"
/translation="MNKNIAIAMSGLTVLASAAVFAADTKVETGDOGTYVQSKYK
KAVEQLGILDSITEIKVFEGTSLASTIKVGSINLAADASKLFTQVDNKLNLGD
GYVDFLIIPSGQGIKITTSLKLVALKDLITAGSADAI IAGTSSADGVVNTNGAASGTE
TNSAGTKLAMSAI FDTAYDSSETA VKITIKADMNDTKFGKAGETTYTSLTFLFDGST
EKIVKLGDSDI1DITKALKLTVVPGSKATVFAEKTGPSAVQPVITKLRINAKETI
DIDASSKTAQDLAKYVFNKTDINTLYKVLNGDEADNTGLIEEVSQKQVVLYPEGK
RVTTLSAASASADENSPVKLTLSKDKKDKDYVDDLTNTYNNYSNAIEVAGEDRIE

TAIALSQKYNSSDDENALFRDSVDNVVLVGGNAIVDGLVASPLASEKAPILLTTSKDK
LDSSVKAEIKRVMNISKSTTGINTSKVYLAGVNSISKVENELKMGKLVTRLAGDD
RYETSLKTADEVGLNDKAFVVGTTGLADAMSIAFPASQLNRANGKMLADGATPIV
VVDGKAKTINDVKDFLDSSQVDIIGGNSVSKDVENAIDDATGKSPDRYSGDDRQAT
NAKVKGSSYYODNLNNDKKVNFVAKDGTGKEDQLVDALAAAPVAANFVGLNSDG
KPVDKGKVLGTGSDNDKNKLVSPAPIVLATDLSLSDQSQSVISKVLDKDNGENLVQVGK
GIATSVINKLKDLSM"
1. .72
/gene="sfpA"
/evidence=experimental
73. .1020
/gene="sfpA"
/product="lower S-layer protein"
/evidence=experimental
1021. .2268
/gene="sfpA"
/product="upper S-layer protein"
/evidence=experimental
sig_peptide
mat_peptide
mat_peptide
ORIGIN
Query Match 29.3%; Score 535.6; DB 1; Length 2271;
Best Local Similarity 59.2%; Pred. No. 7.3e-64;
Matches 1113; Conservative 0; Mismatches 674; Indels 93; Gaps 8;
QY 32 CAGCTGTTACTGTAGTAGTGTCTGCTCCAGTTTTTGGCAGCAGCTTCAGATGTAATAT 91
Db 401 CAGCAGATGCTATATATGCTGGAAACATCTTCAGCAGAGTGGTGTGTTTACAAATACTGGAG 460
QY 92 CACTACAGAGTGGTCAAAATGATAGTATACAGTATCAAACTACTAAAGCTAGTGACTTAG 151
Db 461 CTGCTAGTGGTCTTACTGAGCAAAATCAGCAGGAACAAAACCTTGCAATGTCAGCTATTT 520
QY 152 TAAAGGATATTTTAGCAGCACAAAACCTTAAACAACAGGTGCGAGTTATTTTGAACAAAGATA 211
Db 521 TTGACACAGCATATACAGATTCATCTGAAACCTCGCGTTAAGATTACTATAAAGCAGATA 580
QY 212 CAAAAGTTACT---TTCTATGATGCAAAATGAGAAGATTTCTTCAACTCCAATCGAGATA 268
Db 581 TGAATGATACTAAATTTGGTAAAGCAGGTGAGACAACTTATTTCACTGGGCTTACATTTG 640
QY 269 AAAAGGTTTATTTCAGAACAAAACCTTTAACTACAGCTTAATGGAAATGAAGATATGTAAGA 328
Db 641 AAGATGGGTCTACAGAAAATTTGTTAAATTAGGGCAGATGATATTATAGATATACTA 700
QY 329 CAACTTTAAAA---AATTTAGATGCGAGAGATATGCTTATTATAGATTAACTTTAATAA 385
Db 701 AAGCTCTTAAACTTACTGTTGTTCTGGAAGTAAAGCAACTGTTAAAGTTTGTGAAAAAA 760
QY 386 ATGCTAAACCTGTTG-----AAATTAAGTAGTAGCAGCTAGTGAAA 427
Db 761 CACCAAGTGCAGTGTTCACACAGTAAATAACAAGCTTTAGAATAATAATGCTAAAGAAG 820
QY 428 AAAAGATGTTGTTATCTAGTATGCGAAAAATAGTGCAAAAGATATAGTCTGAAAAATATG 487
Db 821 AAACAATAGATTTGACGCTAGTCTTAGTAAAACAGCACAAGATTTAGCTTAAAAATATG 880
QY 488 TGTTTGAAGACAAAGACTTTAGAAAATGCATAAACCTTATAAATGCCTCAGATTTTCAGTA 547
Db 881 TATTTAATAAACGTGATTTAAATCTCTTTATAAGATTTAAATGAGATGAAGCAGATA 940
QY 548 AAACGTAGTGTATCTATCAAGTAGTCTTTTATCCAAAAGGAAGAGATTACAGGTTTCT 607
Db 941 CTAATGGATTAAATAGAGAAGTTAGTGAAAATATCAAGTAGTTCTTTTATCCAGAAAGAA 1000
QY 608 CAACTTATAGAGCTACAAATTTAT---AATGAAGAACTGCATATGTTATACACCAGTAA 664
Db 1001 AAAGAGTTACAACTAAGAGTGTCTGCAAGGCTTCAATTTGCTGATGAAAATTCACCAGTTA 1060
QY 665 TATTAACTCTAAAAATCTACTAGTAGAGATAATTTTAAAGACTGCAGTAGAAGAGTTACAAA 724
Db 1061 AATTAACTCTTAAAGTCAGATAAGAAGAACTTTAAAGATTATCTGGATGATTTTAAAGAA 1120
QY 725 AATTGAATGCTAGTTATTCTTAATACTACAACTTTTAGCTGGTGATGACAGATAACAACAG 784


```
QY 608 CAACTTATAGAGCTCAAAATTAT---AATGAAGGAACTGCATATGCTAATACACCACTAA 564
Db 1001 AAGAGTTCAACTAAGAGTGCCTCAAGGCTTCAATTGCTGATGAAATTCACCAAGTTA 1060
QY 665 TATTAACCTCAAATCTACTAGTAGAGTAATTTAAAGACTCGAGTAGAAGAGTTACAAA 724
Db 1061 AATTAACCTTAAAGTCAGATAGAGAAAGACTTAAAGATTTATGCGATGATTTAAGNA 1120
QY 725 AATTGAATGCTAGTTATTTCTAATACAACTTTTAGCTGGTGATGACAGAATACAAACAG 784
Db 1121 CATATAAATAATGGATATTCAAATGCTATAGAAAGTAGCAGGAGAGATAGAAATAGAACTG 1180
QY 785 CTATAGAGATAGTAAGAAATATTACATAAATGATGGCGAGAAATCAGATCATTCAGCTG 844
Db 1181 CAATAGCATTAAGTCAAAATATTTAACTCTGATGATGAAAA-----TGCTA 1228
QY 845 ATGTTAAAGAGAATGTTAAAAATGTTGATTAGTAGTGCAAAATGCACCTAGTAGATGGAT 904
Db 1229 TATTTAGAGATTCAGTTGATATAGTATGTTGGAGGAATGCAATAGTTGATGGAC 1288
QY 905 TAGTTGCGGCTCTTTAGCAGCAGAAAAAGATGCTCCACTATTTATTAACCTTCAAAAGATA 964
Db 1289 TTGTAGCTCTCTTTAGCTTCTGAAAAAGAAAGCTCTTTATTTAATTAACCTTCAAAAGATA 1348
QY 965 AATTAGATTCGTCAGTAAATCTGAAATAGAGAGATTTTACACTTAAAACTTCAACAG 1024
Db 1349 AATTAGATTCAGCGCTAAAGCTGAAATTAAGAGAGTTATGAAATATAAGAGTACACAG 1408
QY 1025 AAGTA---ACAGAAAAACAGTTTATATAGCTGGTGGAGTTAATAGTGTATCTAAAGAG 1081
Db 1409 GTATAAATCTTCAAGAAAGTTTATTTAGCTGGTGGAGTTAATCTATATCTTAAGAGAG 1468
QY 1082 TTGTAAACAGAAATAGAAATCAATGGGATTTAAAGTTGAAAGATTTCTCAGTGTATGATAG 1141
Db 1469 TAGAAAAATGAAATTAAGAGATATGGGACTTAAAGTTTACAGATTTAGCAGGAGATGATAG 1528
QY 1142 ATGAACCTCTTTAAATAATAGCAGGTGAATAGGCTTAGATTAATGATAGGCTTATAGTAG 1201
Db 1529 ATGAACCTCTCTTAAATAATAGCTGATGATGATGATGATGATGATGATGATGATGATG 1588
QY 1202 TTGTGGGAACAGGATTTAGCAGATGCCATGAGTATAGCTTCAGTTGCTTCTACTAAATAG 1261
Db 1589 TTGGAGGAACAGGATTTAGCAGATGCCATGAGTATAGCTTCAGTTGCTTCTACTAAATAG 1648
QY 1262 ATGCTAAATGGTGTGTAGATAGAAACAAATGGACATGCTACTCCAAATAGTTGTTGATAG 1321
Db 1649 ATGCTAAATGGTAAATGGATTTAGCTGATGGTGTGCTACACCAATAGTTGTTGATAG 1708
QY 1322 GAAAGCTGATAAATAATCTGATGACTTAGATAGTTTCTTAGGAAGCGCTGATGATAGATA 1381
Db 1709 GAAAGCTGATAAATAATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1768
QY 1382 TAATAGGTGGATTTGCAAGTGTATCTGAAAGATGGAAGAGCTATATCAGATGCTACTG 1441
Db 1769 TAATAGGTGGAGAAACAGTGTATCTAAGATGTTGAAATGCAATAGATGATGATGATGAT 1828
QY 1442 GTPAAGGCGTTTCAAGAGTTAAAGGCGAGATAGACAAGACACTAACTCTGAAAGTTATAA 1501
Db 1829 GTPAATCTCCAGATAGATAGTAGGAGATGATAGACAAGCAACTAATGCAAAAGTTATAA 1888
QY 1502 AAACATATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1561
Db 1889 AAGAATCTCTTTATATCAAGATACTTAAATTAATGATAAAGAGTAGTTAATTTCTTTG 1948
QY 1562 GTGCTTCAAGTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1621
Db 1949 TAGCTAAAGATGGTTCTACTAAGAGATCAATAGTTAGTGTCTTTAGCAGCAGCTCCAG 2008
QY 1622 A-----AGAAAGATCAATTAGTTG 1639
Db 2009 TTGCAGCAAACTTTGGTGTAACTCTTAATTTCTGATGGTAAGCCAGTAGATAGATAAGATG 2068
```

```
QY 1640 ATGCATTAGCAGTAGGAGCTGTTGCTGATATAAA-----CTTGCTCCAGTTGTAT 1690
Db 2069 AAGTATTAACCTGCTTCTGATATGATAAAAATAAATATAGTATCTCCAGCACTATAGTAT 2128
QY 1691 TAGCTACTGATTTCTTTATCTCTGATCAATCGGTTGCTATAGCAAAAGTTGTAGAGAAA 1750
Db 2129 TAGCTACTGATTTCTTTATCTCTGATCAATCGGTTGCTATAGTATAGTAAAGTTCTTGATAAG 2188
QY 1751 AATATCTTAAGATTTTAAACACAGTTGCTCAAGGAATAGCTAATTCAGTTTATAAACAANA 1810
Db 2189 ATAATGGAGAAAACCTTAGTTCAAGTTGGTAAGGTATAGCTACTTCAAGTTATAAACAAT 2248
QY 1811 TGAAAGATTTTATAGATATG 1830
Db 2249 TAAAGATTTTATTAAGTATG 2268

RESULT 9
AX512906 2217 bp DNA linear PAT 03-OCT-2002
LOCUS AX512906 Sequence 6 from Patent WO02062379.
DEFINITION AX512906
ACCESSION AX512906
VERSION AX512906.1 GI:23504065
KEYWORDS Clostridium difficile
SOURCE Clostridium difficile
ORGANISM Clostridium difficile
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.
1
REFERENCE
AUTHORS Doyle R., Kelleher, D., Windle, H.J., Walsh, J.B. and Deirdre, N.E.
TITLE Clostridium difficile vaccine
JOURNAL Patent: WO 02062379-A 6 15-AUG-2002;
THE COLLEGE OF THE HOLY AND UNDIVIDED TRINITY OF QUEEN ELIZABETH (IIE)
FEATURES
source
1. 2217
/organism="Clostridium difficile"
/mol_type="unassigned DNA"
/db_xref="taxon:1496"

ORIGIN
Query Match 27.5%; Score 503.4; DB 6; Length 2217;
Best Local Similarity 59.6%; Pred No. 1.8e-59;
Matches 950; Conservative 0; Mismatches 616; Indels 27; Gaps 5;

QY 250 TCAACTCCAACCTGGAGATAAAAAAGTTTATTTCAGAACAACTTTTAACTACAGCTAATGGA 309
Db 640 TCTGTTCTCTACTACAGGCTTAACATTAACCTGCTGATACAACTGCAACAACAGATGTAAT 699
QY 310 AATGAAGATTTATGTAAGACAACTTTAAAAAATTTAGATGCGAGGAGATATGCTATTATA 369
Db 700 ATTTCTGATGTTATGAGTGCATTTAAATTTTAAATGCTGATACGATAGTGGATTCCCA 759
QY 370 GATTTAACTTATAATAATGCTAAAACTGTTGAAAATTTAAAGTAGTAGCAGCTAGTGAANA 429
Db 760 GCTGTTTCATCAGCTTCTACTCTTAGACCAAGTATAAAGTAATTAATGCAAAAGAGNA 819
QY 430 ACAGTAGTTGATCTAGTAGTCGCAAAAATAGTGCAAAAGATATAGCTGGAATAATATGTG 489
Db 820 TCTATAGATGTTGATTCAAGTTTCACATAGAACAGCTGAAGATTTAGCTGAAAAATATGTA 879
QY 490 TTTGAGACAAGACTTTAGAAAATGCATAAAACTATAAATGCCTCAGATTC----- 543
Db 880 TTTAAACAGAGATGTTGAATAAAAACCTTATGAGGCACCTGATGATTTTATAAAGAGGT 939
QY 544 AGTAAAACTGATGTTACTATCAAGTAGTTCTTTTATCCAAAAGGAAGAGATTACAGGT 603
Db 940 ATACCAAGTATCTTATCCTCAAGATGCTGGAAAATATCAAGTTGTTTATTTGCTCNA 999
QY 604 TTCTCAACTTATAGAGCTACAAAATTAATGAAGGAACTGCAATATGGTAAATACACAGTA 663
Db 1000 GAAAAGAGATTAACCTACTTAAAGGAGCAACTGGAACTTTTAGCAGATGAAAAATTTCTCTCT 1059
```

664 ATATTAACTCTAAATCTACTAGTAAGAGTAATTTAAAGACTGCGAGTAAAGAGTTACAA 723
1060 AAGTAACCAATTAAGCAGATAAAGTAAAGACTTAAAGATATGTTGAGATTTAAA 1119
724 AAATTGAATGCTAGTTATTCTAATCTACTACAACTTTAGCTGGTGAAGACAAATACAA 783
1120 NATGCTAACATGGATATTCAAAATCTGTTGTTAGCAGGTGAAGATAGATAAACA 1179
784 GCTATAGATAGTAAGTAAGAATATTAACAATATATGATGGCGAGAAATCAGATCAATCAGCT 843
1180 GCAATAGAGTTAAGTAGCAAAATATATATACTCTGATGATGACAA-----TGCA 1227
844 GATGTTAAGAGAACTGTTAAATATGTTATAGTAGGTGCAATGCACTAGTAGATGA 903
1228 ATAACCTAAGATCCAGTTAAACATGTTTATAGTTGTTCTCAAGCTGATGATGG 1287
904 TTAGTTGCGGCTCTTTAGCAGCAGAAAGATGCTCCACTATATTAATTAATCTCAAAAGAT 963
1288 CTGTAGCTTCACTTTAGCATCTGAAAGAGAGCTCTTTACTATTAACTTCAGCAGCA 1347
964 AAATTAGATTCGTCAGTAAATCTGAATTAAGAGAGTTTATAGACTTAAACTTCAACA 1023
1348 AAATTAGATTCAGTCTTAAAGCTGATGTTGAAAGAGATTAATGGAATTTAAATCTACA 1407
1024 GAAGTA---ACAGGAAACAGTTTATATAGCTGGTGGAGTTAATAGTGTATCTAAAGAA 1080
1408 GGTGTAATACTTCTAAAGAGTTTACTAGCTGGTGGAGTAACTCTATATCTAAGAT 1467
1081 GTTGTAAACAGATTAAGATCAATGGGATTAAGAGTTTGAAGATTTCTCAGTGATGATGA 1140
1468 GTAGAAATGAATTAAGAGATGAGGACTTAAAGTTTACAAGATTATCAGGAGATGATGA 1527
1141 TATGAACTCTTTTAAATATAGCAGGTGAATAGGCTTAGATATGATGAAGGCTTATGTA 1200
1528 TATGAACTCTTTTAGCTATAGCTGATGAATAGGCTTGTGATATGATGAAGCTTTTGT 1587
1201 GTTGGTGACAGATTTAGCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
1588 GTTGGAGGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1647
1261 GATGTAATGCTGTTGTAGATAGAAACAAATGGACATGCTACTCCAAATAGTTGTGTAGAT 1320
1648 AACTCAATGGAGAACTTGA---CTTAAAGGTGATGCAACTCCAATAGTAGTTGTGAT 1704
1321 GGAAGCTGATATAATATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
1705 GGAAGCTGATATAATATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1764
1381 ATATAGTGTGATTTGCAAGTGTATCTGAAAGATGGAAGATGCAAGATGATGATGATGAT 1440
1765 ATAATAGTGTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1824
1441 GGTAAAGCGGTTACAAGAGTTAAAGCGGACGATAGACAGACACTTCTGAGGTTATA 1500
1825 GGAATATCACTGAGATATAGTGGAGAGATAGACAGCAACAAATGCTAAAGTTATA 1884
1501 AAAACATATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
1885 AAGAAGATGATTTCTTTAAAGATGGAAGTTACAAACTTCTTTGTAGCTAAAGAGGT 1944
1561 GGTGCTTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
1945 TCAATAAAGAGATCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2004
1621 AAGAAGATCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
2005 GGTGATACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2064
1681 CCAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
2065 CCAATGTTTATGACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2124
1741 GTA---GGAGAAAAATATTCTAAGATTTAACAAGTTGCTCAAGGATAGCTAATTC 1797

2125 GTAAATGATGCGCTAATACTAAGAACTAGTCAAGTTGGTAAAGGTATAGCTACTTCA 2184
1798 GTTATAACAAAAATGAAAGATTTATTAGATATG 1830
2185 GTTGTAAATAAATAAAGATTTATTAGATATG 2217
CDI300677 2145 bp DNA linear BCT 04-SBP-2001
Clostridium difficile slpA gene for S-layer protein, strain R7404.
AJ300677
S-layer protein; slpA gene.
Clostridium difficile
Clostridium difficile
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
1
Calabi, E., Ward, S., Wren, B., Paxton, T., Panico, M., Morris, H.,
Dell, A., Dougan, G. and Fairweather, N.
Molecular characterization of the surface layer proteins from
Clostridium difficile
Mol. Microbiol. 40 (5), 1187-1199 (2001)
2 (bases 1 to 2145)
Fairweather, N.F.
Direct Submission
Submitted (19-DEC-2000) Fairweather N.F., Biochemistry, Imperial
College, Exhibition Road, London, SW7 2AY, UNITED KINGDOM
Location/Qualifiers
1. 2145
/organism="Clostridium difficile"
/mol_type="genomic DNA"
/strain="R7404 (ribotype #17)"
/db_xref="taxon:1496"
/country="United Kingdom"
1. 2145
/gene="slpA"
1. 2145
/function="surface layer protein"
/evidence=experimental
/transl_table=1
/product="S-layer protein"
/protein_id="CAC35721.1"
/db_xref="GI:13539184"
/translation="MNKNLAMAATAVTVGSAAPFADSTTPGYTVVKNWKKAVKQ
LDGLKXKTIKTSKVSFNGSVGEVTPASSGAKKADRAAEKLYNLVNTQDVGK
DYVDFVYNLATITKABAEALVTKLQYNDKVLINSATDTVKGWSPQVDSKNV
ANAPLKVSMYTPISAITGSDSGSYIAKPEKTSLLYGTGVDATAGKAITVDTASN
EAPNGKGVTDYKSPKATVQGGTGYTKSGWLKDSADMAATGKIKRVTSKAEISD
VSSYSIENLAKVFPNPKVESEYNAIVALQNDGIESDLVQVNGKYQVTFYPEG
KRLTSSADIADADSPAKTIKANKLKDLYVDDLKTNNYNNVTVAGSDRIET
AELSKYNSDDKNATDDAVNNIVLVGSTIVDGLVASPLASEKTAPLLTSKDKL
DSVSKSEIKRVNMLKSDGTINTSKKYLGGVNSISKOVENELKNMGLKVTRLSGEDR
YVSTLADEIGLDNDKAFVGGTGLADMSIAPVASQLKDGATPIVVVDGKAEIS
DDAKSFLGSDVDIIIGKNSVSKIEISDSATGKTPDRISGDDRQATNAEVLKEDY
FKDGEVNVYFVAKDGGTQDLVLAALAAPIAGRPKESAPAIILATDTLSSDQNVAYS
KAPVKGDTNLVQVGKGIASSVINKMKDLDM"
1. 72
/gene="slpA"
/evidence=experimental
73. 1023
/gene="slpA"
/product="lower S-layer protein"
/evidence=experimental
1024. 2142
/gene="slpA"
sig_peptide
mat_peptide
mat_peptide

Db 904 GCTTATAATGCAATAGTGTGCAATTAACAATAATGATGGAATAGAAATCTGATTTAGTACAATTA 963
Qy 574 CTTTATCCAAAAGAAAGAGATTACAAGGTTTCTCAACTTATAG---AGCTACAATTTAT 630
Db 964 GTTAATGGAATAATCAAGATTATTTTCTATCCAGAAGGAAAGATTGGAATCAAACT 1023
Qy 631 AATGAAGGAAGCTGCATATGTTATATACACCAAGTAAATTAATTAATCTCAAAATCTACTAGTAAG 690
Db 1024 GCAGATATAATAGTGTGATGAGATAGTCCAGCTAAATTAATTAATTAAGCTAATAATTA 1083
Qy 691 AGTAATTTAAAGACTGCGAGTGAAGAGTTACAATAATTAATTAATTAATTAATTAATTAATTAAT 750
Db 1084 AAAGATTTAAAGATTTATGTAGATGATTTAAATAACATACAATAATTAATTAATTAATTAAT 1143
Qy 751 ACAACTTTAGCTGTGATGACAGAAATACAACAGCTATATAGAGTAATTAAGTAATTAATTAAT 810
Db 1144 GTAACAGTAGCAGGAGAGATAGATAGAACTGCTATAGAAATTAAGTAGTAATTAATTAAT 1203
Qy 811 AATAATGATGGCGAGAAATCAGATCAATTCAGCTGATGTTAAAGAGAAATGTTAAATAATGTT 870
Db 1204 AATTCGATGATAAATAATGCAATCACTGATGATGCGATTAAT-----ATATA 1251
Qy 871 GTATTAGTAGTGCATATGCACTAGTAGATGGAATTAGTTGCGGCTCCTTTAGCAGAGAA 930
Db 1252 GTATTAGTTGGATCTACATCTATAGTTGATGGTCTTTGTCATCACCATTAGCTTCAGAA 1311
Qy 931 AAGATGCTCCACTATTTAATTAATTCCTCAAGATTAATTAATTAATTAATTAATTAATTAAT 990
Db 1312 AAAACAGCTCCATTTATTAATTAATTCCTCAAGATTAATTAATTAATTAATTAATTAAT 1371
Qy 991 ATAAGAGAGCTTTAGACTTTAAATCTTCAACAGNAGTA---ACAGAAAAACAGTTTAT 1047
Db 1372 ATAAAAGAGTTATGAATCTTAAGAGTGATCTGTTAATAATTAATTAATTAATTAATTAAT 1431
Qy 1048 ATAGCTGTGGAGTTAATAGTGTATCTAAAGAGTTGTAAACAGAAATTAAGTAATCAATGGGA 1107
Db 1432 TTAGCTGTGGAGTTAATCTTATATCTAAAGATGTAGAAAATGAATTTGAAAAATATGGGC 1491
Qy 1108 TTAAGAGTTGAAGATTTCTCAGGTGATGATAGATTAAGAACTTCTTTAAAAATAGCAGGT 1167
Db 1492 CTTAAAGTTACTAGATTATCAGGAGAAAGACAGATACGAAATCTCTTTTAGCAATAGCTGAT 1551
Qy 1168 GAATAGGCTTAGATATAGTATAGGCTTATGTTAGTGTGCGAAACAGGATTAAGCAGATGCC 1227
Db 1552 GAATAGGCTTGTATATAGTATGAAGCAATTTGTTAGTGTGCTGATCTGATTAAGCAGATGCT 1611
Qy 1228 ATGAGTATAGCTTCAGTTGCTTCTACTATAATTTAGATGGTAATGGTGTGTAGATAGAACA 1287
Db 1612 ATGATATAGCTCCAGTTGCTTCT-----CAACTTTAAA 1644
Qy 1288 AATGGACATGCTACTCCAATAGTTGTTGTAGATGGAAGGCTGATTAATAATCTGATGAC 1347
Db 1645 GATGGAGATGCTACTCCAATAGTATGTTGTAGATGGAAGGCTGATTAATAATGATGATGAT 1704
Qy 1348 TTAGATAGTTTCTTAGAGCGCTGATGATAGATTAATAGTGTGATTTGCAAGTGTATCT 1407
Db 1705 GCTAAGAGTTTCTTAGAGCACTTCTGATGTTGATATAATAGTGTGGAATAATAGCGTATCT 1764
Qy 1408 GMAAGATGGAAGAGCTATATCAGATGCTACTGTTAAAGCGCTTACAAGGTTTAAAGGC 1467
Db 1765 AAAGAGATTGAAGAGTCAATAGATAGTGTCAACTCGGAATACTCCAGATAGATAAGTGA 1824
Qy 1468 GACGATAGACAAGACATTAATCTGGAAGTTTAAATAATTAATTAATTAATTAATTAATTAAT 1527
Db 1825 GATGACAGACAAGCAACTAATGCTGGAAGTTTAAAA----- 1860
Qy 1528 ATAGCTAAAGCTGACGTTTATAGATAAGATTCAGTGCTTCAAGTAGTGTGATGACGAGGTA 1587
Db 1861 -----GAAATGATTAATTTCAAGATGGTGAAGTT 1890
Qy 1588 TTTAATTTCTATGATGCTAAAGATGATCTCAAAAGAAAGATCAATTAATTAATTAATTAATTA 1647
Db 1891 GTGAATTTACTTTTGTGCAAGATGGTCTTACTTAAGAGATCAATTAATTAATTAATTAATTA 1950

Qy 1648 GCAGTAGGAGCTGTGCTGG-----ATATAAATCTGCTCCAGTTGTATTAGCTACT 1698
Db 1951 GCAGCAGCACCATAAGTAGCAGGTAGATTAAAGGAGTCTCCAGCTCCAATCATCTAGCTACT 2010
Qy 1699 GATCTTTATCTTCTGATCAATCGGTGCTATAGCAAGTTGTAGGAGAAAAATTTCT 1758
Db 2011 GATACTTTTCTTCTGACCAAAATGAGTGTGAAGTAAAGCAGTTTCTTAAGATGGTGA 2070
Qy 1759 AAAGATTTAAACAGAGTTGGTCAAGATAGCTAAATTCAGTTTATATAAACAATAAGAT 1818
Db 2071 ACTAACTTAGTTCAAGTAGTAAAGTATAGCTTCTTCAGTTTATAAACAATAAGAT 2130
Qy 1819 TTATTAGATATG 1830
Db 2131 TTATTAGATATG 2142
RESULT 12
AX512907
LOCUS AX512907 2145 bp DNA linear PAT 03-OCT-2002
DEFINITION Sequence 7 from Patent WO02062379.
ACCESSION AX512907
VERSION AX512907.1 GI:23504066
KEYWORDS
SOURCE
ORGANISM
Clostridium difficile
Clostridium difficile
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE
1
AUTHORS Doyle,R., Kelleher,D., Windle,H.J., Walsh,J.B. and Deirdre,N.B.
TITLE Clostridium difficile vaccine
JOURNAL Patent: WO 02062379-A 7 15-AUG-2002;
THE COLLEGE OF THE HOLY AND UNDIVIDED TRINITY OF QUEEN ELIZABETH
(IE)
FEATURES
source
1. .2145
/organism="Clostridium difficile"
/mol_type="unassigned DNA"
/db_xref="taxon:1496"
ORIGIN
Query Match 26.2%; Score 480; DB 6; Length 2145;
Best Local Similarity 62.6%; Pred. No. 2.8e-56;
Matches 909; Conservative 0; Mismatches 435; Indels 108; Gaps 6;
Qy 394 ACTGTTGAAATTAAGATAGTAGCAGCTAGTGAAAAACAGTAGTTGTATCTAGTGCG 453
Db 784 ACTATAAAGTTAGAGTTCAAGTGCAAAAGAAAGAAATCTATTGATGGGATTCAGATTCA 843
Qy 454 AAAAATAGTGCAAAAGATATAGCTGAAAAATATGTGTTTCAAGACAAGAACTTAGAAAAAT 513
Db 844 TATATTAGTGCTGAAAATTTAGCTAAAAAATATGTTAATCTTAAGAGAGTTTCTGAA 903
Qy 514 GCACATAAACTATAAATGCTCAGATTTTCAGTAAAACTGATGTTACTATCAAGTAGTT 573
Db 904 GCTTATAATGCAATAGTTGTCATTACAAAATGATGGAATAGAACTCTGATTAGTACAATTA 963
Qy 574 CTTTATCCAAAAGAAAGAGATTACAAGGTTTCTCAACTTATAG---AGCTACAATAAT 630
Db 964 GTTAATGGAAAAATATCAAGTTATTTTCTATCCAGAAGGAAAAAGATTAGAAACTAAATCT 1023
Qy 631 AATGAAGGACTGCATATGTTAATATACACCAAGTAAATTAATTAATCTTAAAAATCTACTAGTAAG 690
Db 1024 GCAGATATAATAGTGTGATGAGATAGTCCAGCTAAATAATTAATTAAGCTAATAATTA 1083
Qy 691 AGTAATTTAAAGACTGCGAGTGAAGAGTTACAATAATTAATTAATTAATTAATTAATTAAT 750
Db 1084 AAAGATTTAAAGATTTATGTAGATGATTTTAAATAACATACAATAATTAATTAATTAAT 1143
Qy 751 ACAACTTTAGCTGTGATGACAGAAATACAACAGCTATATAGAGTAATTAAGTAATTAATTAAT 810
Db 1144 GTAACAGTAGCAGGAGAGATAGAAATAGAACTGCTATAGAAATTAAGTAGTAATTAATTAAT 1203

Qy	811	AATAATGATGCGGAGAANTCAGATCATCTCAGCTGATGTTTAAAGAGAATGTTAAAAATGTT	870
Db	1204	AAATCTGATGATAAAATGCAATAAATCAGATGATGACGTTAAAT-----AATATA	1251
Qy	871	GTATTAGTGTAGGTGCAAAATGCACCTAGTAGATGGAATTAGTTGCGGCTCCTTTAGCAGAGAA	930
Db	1252	GTATTAGTTGGATCTACATCTATAGTTGATGGTCTTGTTCGATCACATTTAGCTTCAGAA	1311
Qy	931	AAAGATGCTCCACTATTATTAACTTCAAAAGATAAAATAGATTGGTCAGTAAAAATCTGAA	990
Db	1312	AAAACAGCTCCATTTATTATTAACTTCAAAAGATAAAATAGATTCTCAGTAAAAATCTGAG	1371
Qy	991	ATAAAGAGAGTTTTAGACTTTAAAAAATCTCAACAGAGTA--ACAGSAAAAACAGTTTAT	1047
Db	1372	ATAAAAAGAGTTATGAACCTTAAAGAGTGATCTGGTATAAATACCTCTAAAAAAGTTTAT	1431
Qy	1048	ATAGCTGGTGGAGTTAATAGTGTATCTATAAAGAAAGTTGTTAAACAGAAATAGAAATCAATGGGA	1107
Db	1432	TTAGCTGTTGGAGTTAATCTATATCTAAAGATGTAGAAGATGAAATTGAAAAATATGGGC	1491
Qy	1108	TTAAAAGTTGAAGAATCTCAGGTGATGATAGATATGAAAACTTCTTTAAAAAATAGCAGGT	1167
Db	1492	CTTAAAGTTTACTAGATTATCAGGAGAAGACAGATACGAAACTCTCTTTAGCAATAGCTGAT	1551
Qy	1168	GAAATAGCTTTAGATAATGATTAAGGCTTTATGTAGTTGGTGGAAACAGGATTTAGCAGATGCC	1227
Db	1552	GAAATAGCTTTGATAATGATTAAGCAATTTGTAGTTGGTGGTACTGGAATTCGCAGATGCT	1611
Qy	1228	ATGAGTATAGCTTCAGTTGCTTCTACTATAAATTAGANGTTAATGGTGTGTGTAGATAGAACA	1287
Db	1612	ATGAGTATAGCTCCAGTTGCTTCT-----CAACTTAAA	1644
Qy	1288	AATGGAATGCTACTCCAATAGTTGTTGTAGATGGAAGAAAGCTGATAAAATATCTGATGAC	1347
Db	1645	GATGGAGATGCTACTCCAATAGTAGTTGTGTAGATGGAAGAAAGCAAAAGAAATTAAGTGTAT	1704
Qy	1348	TTGATATAGTTTCTTAGGAAGCGCTGATGTAGATATATAGTGGATTTGCCAAGTGTATCT	1407
Db	1705	GCTAAGAGTTTCTTAGGAACTTCTGATGTTGATATANTAGTGGAAAAAATAGCTATCT	1764
Qy	1408	GAAAAGATGGAAGAGCTATATCAGATGCTACTGTTAAAGCGGTTACAAGATTAAAGGC	1467
Db	1765	AAAGAGATTTGAAGAGTCAATAGATAGTGCACCTGGAAAAAACTCCAGATAGATAAAGTGA	1824
Qy	1468	GACGATAGACAAGACACTAACTCTGGAAGTTTATAAAACATATTATGCTAATGATCTGAA	1527
Db	1825	GATGACAGACAGCACTATGCTGAAGTTTAAAA-----1860	
Qy	1528	ATAGCTAAAGCTGCAGTTTGTAGATAAAGATTCAGGTGCTTCAAAGTAGTGATGAGGATTA	1587
Db	1861	-----GAAGATGATTATTTCAAAGATGGTGAAGTT	1890
Qy	1588	TTTTAATTTCTATGTAGCTAAGATGGATCTACAAAAGAAGATCAATTAGTTGATGCAATTA	1647
Db	1891	GTGAATTTACTTTGTGCAAAAGATGGTTCTACTAAAGAAGATCAATTTAGTAGATGCAATTA	1950
Qy	1648	GCAGTAGAGCTGTGCTGG-----ATATAAACTGCTCCAGTTGTATTAGCTACT	1698
Db	1951	GCAGCAGCAACCAATAGCAGGTAGATTTAAGGAGTCTCCAGCTCCAATCATACTAGCTACT	2010
Qy	1699	GATTCTTTATCTTGATCAATCGGTTGCTTATAAGCAAAAGTTGTAGGAGAAAAATATTCT	1758
Db	2011	GATACTTTATCTTCTGACCAAAATGTAGCTGTGTAGTAAAGCAGTTCTCAAAGATGGTGA	2070
Qy	1759	AAAGATTTAAACACAAGTTGGTCAAGGAATAGCTAATTCAGTTTATAACAAAAATGAAGAT	1818
Db	2071	ACTAACTTAGTTCAAGTAGGTAAAGGTATAGCTTCTTCAGTTTATAACAAAAATGAAGAT	2130
Qy	1819	TTATTAGATATG	1830
Db	2131	TTATTAGATATG	2142

956 CAAAGAGTAATAGATTGCTGAGTAAATCTGAAATAAAGAGAGTGTAGACTTAAAAA 1015
 1352 CAAAGATTAATTTAGATTCTAGTAAATCTGAAATAAAGAGAGTGTAGACTTAAAGA 1411
 1016 CTTCAACAGAGTA---ACAGGAAACAGTGTATATAGCTGGTGGAGTAAATAGTAT 1072
 1412 GTGACACTGGTATAAATCTCTTAAAGAGTGTATATAGCTGGTGGAGTAAATCTATAT 1471
 1073 CTAAGAGTGTAAACAGATTAAGATCAATGAGTAAAGTGTAAAGTGTAAAGTGTAAAG 1132
 1472 CTAAGAGTGTAAAGATTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAG 1531
 1133 ATGATAGATTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAG 1192
 1532 AAGACAGATTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAG 1591
 1193 CTTATGATGTTGGTGAACAGATTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGT 1252
 1592 CATTGATGTTGGTGAACAGATTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGT 1650
 1253 CTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAG 1312
 1651 CTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAG 1684
 1313 TTGTAGATGAAGAGCTGATAAATATCTGATGATGATGATGATGATGATGATGATGAT 1372
 1685 TTGTAGATGAAGAGCTGATAAATATCTGATGATGATGATGATGATGATGATGATGAT 1744
 1373 ATGTAGATTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAG 1432
 1745 ATGTAGATTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAG 1804
 1433 ATGTAGATTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAG 1492
 1805 GTGCACTGCAAACTCCAGATGAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGT 1864
 1493 AAGTTATGAAGATTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGT 1552
 1865 AAGTTATGAAGATTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGT 1890
 1553 AAGATTGAGTGTCTCAAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGT 1612
 1891 AAGATTGAGTGTCTCAAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGT 1930
 1613 GATCTCAAAAGAGATCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1667
 1931 GTTCTCAAAAGAGATCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1990
 1668 ATATAAAGTGTCTCAAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGT 1723
 1991 TTAAGAGTGTCTCAAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 2050
 1724 TTGCTATAAGCAAAAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 1783
 2051 TAGCTGTAAAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 2110
 1784 GAATAGCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1830
 2111 GTATAGCTTCTCAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAG 2157

RESULT 14
 AX512903
 LOCUS
 DEFINITION Sequence 3 from Patent WO02062379.
 AX512903
 ACCESSION
 VERSION AX512903.1 GI:23504062
 KEYWORDS
 SOURCE
 ORGANISM
 Clostridium difficile
 Clostridium difficile
 Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 Clostridium.
 REFERENCE
 1

AUTHORS Doyle, R., Kelleher, D., Windle, H.J., Walsh, J.B. and Deirdre, N.E.
 TITLE Clostridium difficile vaccine
 JOURNAL Patent: WO 02062379-A 3 15-AUG-2002;
 THE COLLEGE OF THE HOLY AND UNDIVIDED TRINITY OF QUEEN ELIZABETH
 (IE)
 FEATURES Location/Qualifiers
 source
 1..2158
 /organism="Clostridium difficile"
 /mol_type="unassigned DNA"
 /db_xref="taxon:1496"
 ORIGIN
 Query Match 25.8%; Score 471.8; DB 6; Length 2158;
 Best Local Similarity 61.5%; Pred. No. 3.6e-55;
 Matches 951; Conservative 0; Mismatches 467; Indels 129; Gaps 7;
 296 CTACAGCTAATGGAATGAAGATTATGTAAGACAACTTTAAAGATTTAGATGACAGGAG 355
 728 CAACACCTAGTCTGTAGCTGTAAGTGGTTTGTACTAAAGATGATGATCTGATTTAGCAA 787
 356 AATATGCTATTATAGATTAACTTATATAAATGCTAAAACTGTGAAATTTAAAGTAGTAG 415
 788 AATCAGTACTATTAATGTAAGATTATTAATGCAAAAGAGAAATCAATTTGATATAGATG 847
 416 CAGCTAGTGAAGAAACAGATGTTGTATCTAGTATGCGAAAGTAAAGTGTGCAAAAGATATAG 475
 848 CAAGCTCATATACATCAGCTG-----AAAATTTAGCTAAAGATATGTTTGTATCCAG 901
 476 CTGAAAAATATGTTGTTGAAGCAAGACTTAAAGATGCACCTTAAAGATCTATAAATGCCT 535
 902 ATGAATTTCTGAAGCATATAAGCAATAGTATGATTAACAAATGATGGTATAGATGCTTA 961
 536 CAGATTTTCAGTAAACCTGATAGTTACTATCAAGTACTGTTCTTATCCAAAGGAAAGAGAT 595
 962 ACTTAGTTCAGTTAGTTAATGAAATATCAAGTATTTTATCCAGAGGTAAAGAT 1021
 596 TACAAGTCTTCAACTTATAGAGCTACAAATTTAATGAAGAACTGCAATATGGTAAATA 655
 1022 TAGAACTAAATCAGCA-----AATGATACAATAGCTAGTCAAGATA 1063
 656 CACAGTAAATTAACCTTAAATCTACTAGTAAAGATTAATTTAAAGCTGCAAGTAGAAG 715
 1064 CACAGCTAAGTATGTTAATAAGAGCTAAATAATTTAAAGATTTAAAGATTTATGTAGATG 1123
 716 AGTTACAAAAATGGAATGCTAGTTATTTCTAATACTACAACTTTAGCTGGTGTGACAGAA 775
 1124 ATTTAAACATATAATATATCTTTTCAATGTTGTACAGTACAGGAGAGATAGAA 1183
 776 TACAAACAGCTATAGAGATAAGTAAAGATATTACAAATATATATGATGGGAGAAATCAGATC 835
 1184 TAGAACTGCTATAGAAATTAAGTAGTAAATATTATAATTTCTGAT-----GATA 1231
 836 ATTCAGCTGATGTTAAGAGAAATTTAAAGATCTTTTATTTAGTGGTCAATGCACATAG 895
 1232 AAAATGCAATTAACGTAAAGCAGTTAATGATATAGTATTAGTTGGATCTACATCTATAG 1291
 896 TAGATGATTTAGTTGGCTCTTTTAGCAGCAGAAAAAGATGCTCCACTATTATTAATCTT 955
 1292 TTGATGCTTTGTTGCAATCACCATTGCTTCAGAAAAACAGCTCCATTATTATTAATCTT 1351
 956 CAAAGATAAATTTAGATTTCTGATAAATCTGAAATTAAGAGAGTGTAGACTTAAAGAA 1015
 1352 CAAAGATAAATTTAGATTTCTGATAAATCTGAAATTAAGAGAGTGTATGAATCTAAAGA 1411
 1016 CTTCAACAGAGTA---ACAGGAAACAGTTTATATAGCTGTTGGAGTAAATAGTCTAT 1072
 1412 GTGACACTGGTATAAATCTTCTAAAAAGTGTATTTAGCTGGTGGAGTTAAATTTCTATAT 1471
 1073 CTAAGAGTGTAAACAGATTAAGTAAATCAATGGAATTTAAAGTGTAAAGTGTAAAGTGT 1132
 1472 CTAAGAGTGTAAAGTGTAAATGGAATTTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAG 1531
 1133 ATGATAGATTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAG 1192

Db 1532 AAGACAGATCGAAGACTCTTTAGCATTAGCTGATGAATAGTCTTGATTAATGATAAG 1591
QY 1193 CTTATGTAGTTGGTGGAAAGCAGGATTAGCAGATGCCATGATAGTCTCAGTTGCTTCTTA 1252
Db 1592 CATTTGTAGTTGGTGGTACTGCGATTAGCAGATGCTATGATGATAGTCTCAGTTGCTTCT- 1650
QY 1253 CTAAATTAGATGGTAATGCTGTTGTAGATAGAACAAATGGACATGCTACTCCAAATAGTTG 1312
Db 1651 -----CAACTTAAAGATGGAGATGCTACTCCAAATAGTAG 1684
QY 1313 TTGTAGATGAAAAGCTGATAAAATATCTGATGACTTATAGATAGTTCTTTAGGAAGCGCTG 1372
Db 1685 TTGTAGATGAAAAGCAAGAAATAGTATGATGATGCTTAAGAGTTCTTTAGGAAGCTTCTG 1744
QY 1373 ATGTAGATATAATAGTGGATTGTCAGAGTGTATCTGAAAAGATGGAAGAGCTATATACAG 1432
Db 1745 ATGTTGATATAATAGTGGAAAATAATAGGTAATCTAAAGAGATTGAAGAGTCAATAGATA 1804
QY 1433 ATGCTACTGGTAAAGCGTTACAAGAGTTAAAGGCGACGATAGACAGACACTAACTCTG 1492
Db 1805 GTGCAACTGGAAAACCTCCAGATAGAATAAGTGGAGATGATAGACAAGCAACTAAATGCTG 1864
QY 1493 AAGTTATAAAAAACATATTATGCTTAATGATGACTGAAATAGCTAAAGCTGCAGTTTATAGATA 1552
Db 1865 AAGTTTAAAGAGATGATTAATTC----- 1890
QY 1553 AAGATTCAAGTGTCTCAAGTAGTATGATGCGAGGATATTTAATTTCTATGTAGCTAAAGATG 1612
Db 1891 -----ACAGATGGTGAAGTTGTGAATTACTTTTGTGCAAAAGATG 1930
QY 1613 GATCTACAAAAGAGATCAATTAGTTGATGCAATTAGCAGTAGGAGCTGTGCTGG----- 1667
Db 1931 GTTCTACTAAAGAGATCAATTAGTAGATGCGCTTAGCAGCAGCACCAATAGCAGGTAGAT 1990
QY 1668 ----ATATAAATCTGCTCCAGTTGTATTAGTACTGATTTCTTTATCTCTGATCAATCGG 1723
Db 1991 TTAAGAGTCTCAGCTCCATCATCTACTAGTACTGATGATCTTTATCTCTGACCAAAATG 2050
QY 1724 TTGCTATAAGCAAGTTGTAGGAGAAAATAATTTCTAAAGATTTAAACAAAGTTGTCAG 1783
Db 2051 TAGCTGTAAGTAAAGCAGTTCTCTAAAGATGTTGGAACCTAACTTAGTTCAAGTAGGTAAAG 2110
QY 1784 GAATAGCTAATTCAGTTATAACCAAAATCAAGATTTTATAGATATG 1830
Db 2111 GTATAGCTTCTTCAGTTATAACAAATGAAGATTTTATTAGATATG 2157

RESULT 15
AX512908
LOCUS AX512908 2158 bp DNA linear PAT 03-OCT-2002
DEFINITION Sequence 8 from Patent WO02062379.
ACCESSION AX512908
VERSION AX512908.1 GI:23504067
KEYWORDS Clostridium difficile
SOURCE Clostridium difficile
ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.
REFERENCE 1
AUTHORS Doyle, R., Kelleher, D., Windle, H. J., Walsh, J. B. and Deirdre, N. E.
TITLE Clostridium difficile vaccine
JOURNAL Patent: WO 02062379-A 8 15-AUG-2002;
THE COLLEGE OF THE HOLY AND UNDIVIDED TRINITY OF QUEEN ELIZABETH (IE)
FEATURES Location/Qualifiers
1..2158
/organism="Clostridium difficile"
/mol_type="unassigned DNA"
/db_xref="taxon:1496"
ORIGIN
Query Match 25.8%; Score 471.8; DB 6; Length 2158;

Best Local Similarity 61.5%; Pred. No. 3.6e-55;
Matches 951; Conservative 0; Mismatches 467; Indels 129; Gaps 7;
QY 296 CTACAGCTAAATGGAAATGAAGATTATGTAAGACAACTTTTAAAAAATTTAGATGCGAGAG 355
Db 728 CAACACCTAGTGTCTGTAGCTGTAGAGTGTGTAACCTAAAGATGATGACTGATTTAGCAA 787
QY 356 AATATGCTTATATAGATTAACTTATATATATGCTTAAACTGTTGAAATTTAAAGTAGTAG 415
Db 788 AATCAGGTACTATTAATGTAAAGTTATAAATGCAAAAAGAAATCAATTTGATATAGATG 847
QY 416 CAGCTAGTGA AAAACACGCTAGTTGTATCTAGTGTGCGAAAAATAGTGC AAAAGATATAG 475
Db 848 CAAGCTCATATACATCAGCTG-----AAAATTTAGCTTAAAGACATGTTTGTATCCAG 901
QY 476 CTGAAAAATATGTGTTTGAAGACAAAGACTTTAGAAAAATGCACTAAAACTATAAATGCGT 535
Db 902 ATGAAATTTCTGAAGCATATAAGGCAATAGTAGCATTA CA AAAATGATGTTATAGAGTCTA 961
QY 536 CAGATTTCAAGTAAAACTGATAGTTACTATCAAGTAGTTCTTTATCCAAAAGAAAGAGAT 595
Db 962 ATTTAGTTCAAGTTAGTTAATGGAATAATCAAGTGATTTTTTATCCGAAAGGTA AAAAGAT 1021
QY 596 TACAAGGTTTCTCAACTTTATAGAGCTACAAATTTATAATGAAGAACTGCATATGGTAATA 655
Db 1022 TAGAACTTAATCAGCA-----AATGATACATATAGCTAGTCAAGATA 1063
QY 656 CACCACTAATATTAATCTCAAAATCTACTAGTAAAGATPAATTTAAAGACTGCAGTAGAAG 715
Db 1064 CACCAGCTAAAGTAGTTATAAAGCTAATAAATTA AAAAGATTTAAAGAGATTATGTAGATG 1123
QY 716 AGTTACAAAATTTGATGCTAGTTATTTCTTAATCTACTACAACTTTAGCTGGTGTGACAGA 775
Db 1124 ATTTAAAAACATATATAATTAATCTTATTTCAAAATGTTGTAACAGTAGCAGGAGAAAGATA 1183
QY 776 TACAAACAGCTATAGAGATAAGTAAAGAAATATTAACAATAATGATGGCGAGAAATCAGATC 835
Db 1184 TAGAACTGCTATAGAAATTAAGTAGTAATAATTTATTAATTTCTGAT-----GATA 1231
QY 836 ATTCAGCTGATGTTTAAAGAGAAATGTTAAAAATGTTGTTATTTAGTAGGTGCAATGCAGTAG 895
Db 1232 AAAATGCAATAACTGATAAAGCAGTTAATGATATAGTATTAGTTGGATCTACATCTATAG 1291
QY 896 TAGATGGATTAGTTGGCGCTCTTTAGCAGCAGAGAAAAGATGCTCCACTATTTAATTAACCT 955
Db 1292 TTGATGGTCTTGTGTCATCACCATTTAGCTTCAGAAAAACAGCTCCATTTATTTAATTAACCT 1351
QY 956 CAAAAGATAAAATTTAGATTGCTCAGTAAAAATCTGAAATAAAGAGAGTTTTTAGACTTAAAAA 1015
Db 1352 CAAAAGATAAATTAGATTTCATCAGTAAATCTGAAATAAAGAGAGTTATGAATTAAGA 1411
QY 1016 CTTCAACAGAAAGTA---ACAGGAAAAACAGTTTATATAGCTGGTGGAGTTAATAGTGTAT 1072
Db 1412 GTGACACTGGTATAAATAATCTTCTAAAAAAGTTTATTTAGCTGGTGGAGTTAATTTCTATAT 1471
QY 1073 CTAAAGAGTTGTAAACAGAAATTAGAATCAATGGGATTA AAAAGTTGAAAGATTTCTCAGGTG 1132
Db 1472 CTTAAAGATGTAGAAAATGAATTTGAAAAACATGGGGCTTTAAAAGTTTACTAGATTATCAGGAG 1531
QY 1133 ATGATAGATATGAAAACCTCTTTTAAAAATAGCAGGTTGAAATAGGCTTAGATTAATGATAAGG 1192
Db 1532 AAGACAGATACGAAACTCTTTTAGCAATAGCTGATGAATAGGTTCTTGATTAATGATAAG 1591
QY 1193 CTTATGTAGTTGGTGGAAACAGGATTAGCAGATGCCATGCGATAGCTTCAAGTTGCTTCTTA 1252
Db 1592 CATTTGTAGTTGGTGTACTGCAATTAGCAGATGCTATGATGATAGCTCCAGTTGCTTCT- 1650
QY 1253 CTAAATTTAGATGGTAAATGTTGTTAGATAGAACAAATGGACATGCTACTCCAAATAGTTG 1312
Db 1651 -----CAACTTAAAGATGGAGATGCTACTCCAAATAGTAG 1684
QY 1313 TTGTAGATGAAAAGCTGATAAAATATCTGATGACTTATAGATAGTTCTTTAGGAAGCGCTG 1372

Db 1685 TTCTAGATGAAAGCAAAAGAAATAAGTGATGATGCTAAGAGTTTCTTAGGAACTTCTG 1744
Qy 1373 ATCTAGATATAATAGTGGATTTGCAAGTGTATCTGAAAGATGGAAGAGCTATATATCAG 1432
Db 1745 ATGTTGATATAATAGTGGAAAAAATAGCGTATCTAAAGAGATTGAAGAGTCAATAGATA 1804
Qy 1433 ATGCTACTGTAAAGCGTTTACAAGAGTTTAAAGCGACGATAGACAAGACACTAACTCTG 1492
Db 1805 GTGCAACTGGAAGAACTCCAGATAGATAAGTGGAGATGATAGACAAGCAACTAATGCTG 1864
Qy 1493 AAGTTATAAAACATATATATGCTTAATGATATCTGAAATAGCTAAAGCTGCAGTTTTAGATA 1552
Db 1865 AAGTTTTAAAGAAAGATGATTATTC----- 1890
Qy 1553 AAGATTCAGGTGCTTCNAGTAGTGTGATGCAGGAGTATTTAATTTCTATGTAGCTAAAGATG 1612
Db 1891 -----ACAGATGGTGAAGTTGTGAATTACTTTTGTGCAAAAGATG 1930
Qy 1613 GATCTACAAAAGAGATCAATTAGTTGATGCATTAGCAGTAGGAGCTGTTGCTGG----- 1667
Db 1931 GTTCTACTAAAGAGATCAATTAGTAGATGCTTAGCAGCAGCACCACCAATAGCAGGTAGAT 1990
Qy 1668 ----ATATAAACTTGCTCCAGTTGTATTAGTACTGATTTCTTATCTTCTGATCAATCGG 1723
Db 1991 TTAAGGAGTCTCCAGCTCCAATCATACTAGTACTGATATCTTTATCTTCTGACCAAAATG 2050
Qy 1724 TTGCTATAAGCAAGTTGTAGGAGAAAATATTTCTAAAGATTTAAACACAGTTGGTCAAG 1783
Db 2051 TAGCTGTAAGTAAAGCAGTTTCCTAAAGATGGTGGAACTTAAGTTCAAGTAGGTAAAG 2110
Qy 1784 GAATAGCTAATTCAGTTTATAACAAAATGAAAGATTTATTAGATATG 1830
Db 2111 GTATAGCTTCTTCAGTTATATAACAAAATGAAAGATTTATTAGATATG 2157

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 6, 2005, 04:50:47 ; Search time 1014 Seconds
(without alignments)
10683.557 Million cell updates/sec

Title: US-10-068-870A-4
Perfect score: 1830
Sequence: 1 atgaaaaaaagaatttagc.....tgaaagattattagatatg 1830

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues 8780412
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1980s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1830	100.0	1830	6	ABs65376 Clostridi
2	1830	100.0	1830	6	ABs65368 Clostridi
3	542	29.6	2268	6	ABs65367 Clostridi
4	535.6	29.3	2271	6	ABs65378 Clostridi
5	503.4	27.5	2271	4	AAI65841 Nucleotid
6	503.4	27.5	2217	6	ABs65372 Clostridi
7	503.4	27.5	2217	6	ABs65380 Clostridi
8	483.2	26.4	2145	4	AAI65839 Nucleotid
9	480	26.2	2145	6	ABs65371 Clostridi
10	480	26.2	2145	6	ABs65381 Clostridi
11	471.8	25.8	2157	6	ABs65369 Clostridi
12	471.8	25.8	2157	6	ABs65374 Clostridi
13	471.8	25.8	2157	6	ABs65373 Clostridi
14	471.8	25.8	2157	6	ABs65375 Clostridi
15	471.8	25.8	2158	6	ABs65377 Clostridi
16	471.8	25.8	2158	6	ABs65382 Clostridi
17	471.8	25.8	2160	4	AAI65840 Nucleotid
18	471.8	25.8	48551	6	AAI65840 Nucleotid
19	470.2	25.7	2157	6	ABs65370 Clostridi
20	470.2	25.7	2158	6	ABs65379 Clostridi

21	145.2	7.9	1893	8	ACA28989	Prokaryot
22	112.4	6.1	3399	2	AAT05868	Chicken 1
23	104	5.7	1503	8	ACA28912	Prokaryot
24	99.6	5.4	1830	8	ACA28975	Prokaryot
25	95.8	5.2	1686	2	AAQ87587	DNA encod
26	95.6	5.2	5361	2	AAT78868	P. falcip
27	95.6	5.2	5529	6	ABK50883	DNA encod
28	95.6	5.2	6152	2	AAT78867	P. falcip
29	92	5.0	4997	8	ACA47636	Prokaryot
30	90.2	4.9	3095	2	AAQ03875	Sequence
31	90.2	4.9	3579	3	AAA70099	Plasmodi
32	88.4	4.8	5253	8	ACA29579	Prokaryot
33	87.6	4.8	7442	4	AA846866	Tumour su
34	86.2	4.7	7446	8	ACA19708	Prokaryot
35	86.2	4.7	7446	10	ACF05848	Methicill
36	85.2	4.7	4965	8	ADA89806	Staphyloc
37	85.2	4.7	7434	4	AA552179	Staphyloc
38	85.2	4.7	7434	8	ABT14965	Pathogen
39	85.2	4.7	7437	4	ACF73459	Staphyloc
40	85.2	4.7	7437	4	AA555232	Staphyloc
41	85.2	4.7	8155	2	AAV74374	Staphyloc
42	84.8	4.6	1300	12	ADP85917	Synthetic
43	84.2	4.6	4911	8	ABT15076	Pathogen
44	82.2	4.5	2370	4	AAH28330	Nucleotid
45	81.8	4.5	6292	4	AA846735	Tumour su

ALIGNMENTS

RESULT 1

ABs65376
ID ABs65376 standard; DNA; 1830 BP.

AC ABs65376;

DT 15-NOV-2002 (first entry)

DE Clostridium difficile strain 172450 slpA DNA sequence.

XX Vaccine; Clostridium difficile; immunogenic; immunoreactivity;
KW C. difficile infection; immune response; C. difficile associated disease;
KW passive immunotherapy; passive vaccination; interleukin; adjuvant;
KW C. difficile vaccine; antibacterial; slpA; ds.

OS Clostridium difficile.

PN WO200262379-A2.

PD 15-AUG-2002.

PF 11-FEB-2002; 2002WO-1E000017.

PR 09-FEB-2001; 2001IE-00000137.

PA (QUEB-) QUEEN ELIZABETH COLLEGE DUBLIN.

PI Doyle R, Kelleher D, Windle HJ, Walsh JB, Deirdre NE;

DR WPI; 2002-643389/69.

PT Novel vaccine useful for treatment/prophylaxis of Clostridium difficile associated diseases, comprises Clostridium difficile genes or peptide/polypeptides or its derivative, fragment, mutant or variant.

PS Claim 12; Page 70-71; 85pp; English.

XX The present invention relates to a new vaccine for treatment/prophylaxis of Clostridium difficile associated disease. The vaccine of the invention comprises C. difficile gene or peptide/polypeptide or derivative, fragment, mutant or variant of the peptide/polypeptide which is immunogenic in humans or to which immunoreactivity is detected in individuals who have recovered from C. difficile infection. The molecules

CC of the invention are used in prophylaxis or treatment of C. difficile
 CC associated disease, where the medicament is a vaccine. The invention can
 CC be used by administering the vaccine preparation to a host to raise an
 CC immune response. The invention is further useful in the preparation of a
 CC medicament for treatment or prophylaxis of C. difficile infection or C.
 CC difficile associated disease. The invention is also useful in passive
 CC immunotherapy for established C. difficile infection, for the eradication
 CC of C. difficile associated disease, or for passive vaccination of an
 CC individual with C. difficile infection. Interleukin is useful as an
 CC adjuvant in C. difficile vaccine. The present nucleic acid sequence
 CC represents a Clostridium difficile slpA DNA sequence of the invention
 XX
 SQ Sequence 1830 BP; 716 A; 220 C; 362 G; 532 T; 0 U; 0 Other;

Query Match		100.0%;	Score 1830;	DB 6;	Length 1830;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 1830;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGMAAAAGAAATTTAGCAATGGCTATGGCAGCTGTTACTGTAGTAGTTCCTGCTGCT	60		
DB	1	ATGMAAAAGAAATTTAGCAATGGCTATGGCAGCTGTTACTGTAGTAGTTCCTGCTGCT	60		
QY	61	CCAGTTTTTGCAGCAGCTTCAGATGTAATATCACTACAAGATGGTACAAATGATAAT	120		
DB	61	CCAGTTTTTGCAGCAGCTTCAGATGTAATATCACTACAAGATGGTACAAATGATAAT	120		
QY	121	ACAGTATCAAACTAAGCTAGTGACTTAGTAAGAGATATTTTAGCAGCACAAACTT	180		
DB	121	ACAGTATCAAACTAAGCTAGTGACTTAGTAAGAGATATTTTAGCAGCACAAACTT	180		
QY	181	ACAAAGGTCAGTATTTTGAACAAAGATACAAAGTTTACTTCTATGATGCAATGAG	240		
DB	181	ACAAAGGTCAGTATTTTGAACAAAGATACAAAGTTTACTTCTATGATGCAATGAG	240		
QY	241	AAAGATCTTCAACTCCAACTGGAGATAAAAAGTTTATTCAGAACAACTTTAACTACA	300		
DB	241	AAAGATCTTCAACTCCAACTGGAGATAAAAAGTTTATTCAGAACAACTTTAACTACA	300		
QY	301	GCTAATGGAATGAAGATTATGTAAGACAACTTTTAAAGATTTAGATGACAGAGATAT	360		
DB	301	GCTAATGGAATGAAGATTATGTAAGACAACTTTTAAAGATTTAGATGACAGAGATAT	360		
QY	361	GCTATTATAGATTAACTTATTAATATGCTAAACCTGTTGAAATTTAAAGTAGTACGCT	420		
DB	361	GCTATTATAGATTAACTTATTAATATGCTAAACCTGTTGAAATTTAAAGTAGTACGCT	420		
QY	421	AGTGAATAAACAAGTATGTTATCTAGTGATGCGAAAAATAGTGCAGAAATATAGCTGAA	480		
DB	421	AGTGAATAAACAAGTATGTTATCTAGTGATGCGAAAAATAGTGCAGAAATATAGCTGAA	480		
QY	481	AAATATGTTTGAAGACAAAGCTTAGAATATGCACTAAAGCTATTAATGCCCTCAGAT	540		
DB	481	AAATATGTTTGAAGACAAAGCTTAGAATATGCACTAAAGCTATTAATGCCCTCAGAT	540		
QY	541	TTCAAGTAAACCTGATGTTACTATCAAGTAGTTCCTTATCCAAAGAGAAAGATATCA	600		
DB	541	TTCAAGTAAACCTGATGTTACTATCAAGTAGTTCCTTATCCAAAGAGAAAGATATCA	600		
QY	601	GGTTTCTCAACTTATAGAGCTACAAATATTAATGAAGAACTGCAATATGTTAATACCA	660		
DB	601	GGTTTCTCAACTTATAGAGCTACAAATATTAATGAAGAACTGCAATATGTTAATACCA	660		
QY	661	GTAATATTAACCTCTAAATCTAGTAAGAGTAATTTAAAGATGCGAGTAGAAGATTA	720		
DB	661	GTAATATTAACCTCTAAATCTAGTAAGAGTAATTTAAAGATGCGAGTAGAAGATTA	720		
QY	721	CAAAATTTGAATGCTAGTTATTTCTAATCTCAACTTATGCTGGTGAACAGATACAA	780		
DB	721	CAAAATTTGAATGCTAGTTATTTCTAATCTCAACTTATGCTGGTGAACAGATACAA	780		
QY	781	ACAGCTATAGAGATAAGTAAAGAAATTTACAAATATGATGGCGAGAAATCAGATCATTC	840		
DB	781	ACAGCTATAGAGATAAGTAAAGAAATTTACAAATATGATGGCGAGAAATCAGATCATTC	840		

RESULT 2
 ABS65368
 ID ABS65368 standard; DNA; 1830 BP.

QY	841	GCTGATGTTAAAGAGAAATGTTTAAAAATGTTGTTATTTAGTAGTGGCAAAATGCATAGTAGAT	900
DB	841	GCTGATGTTTAAAGAGAAATGTTTAAAAATGTTGTTATTTAGTAGTGGCAAAATGCATAGTAGAT	900
QY	901	GGATTAGTTGGCGCTCCCTTTAGCAGCAGAAAAGATGCTCCCACTATTATTAACTTCARAA	960
DB	901	GGATTAGTTGGCGCTCCCTTTAGCAGCAGAAAAGATGCTCCCACTATTATTAACTTCARAA	960
QY	961	GATAAATTAGATTTCGTTCAGTAAAAATCTGAAATTAAGAGAGTTTATAGACTTTAAACCTTCA	1020
DB	961	GATAAATTAGATTTCGTTCAGTAAAAATCTGAAATTAAGAGAGTTTATAGACTTTAAACCTTCA	1020
QY	1021	ACAGAAAGTAAACAGGAAAACAGTTTATATAGCTGCTGGAGTTTAAAGAGTTTAAAGAACTTCA	1080
DB	1021	ACAGAAAGTAAACAGGAAAACAGTTTATATAGCTGCTGGAGTTTAAAGAGTTTAAAGAACTTCA	1080
QY	1081	GTTGTAAACAGAAATTAAGAAATCAATGGATTTAAAGTTTGAAGATTTCTCAGGTGATGATAGA	1140
DB	1081	GTTGTAAACAGAAATTAAGAAATCAATGGATTTAAAGTTTGAAGATTTCTCAGGTGATGATAGA	1140
QY	1141	TATGAAACTTCTTTAAAAATAGCAGTGAATAGGCTTAGATAATATGATAAGGCTTTATGTA	1200
DB	1141	TATGAAACTTCTTTAAAAATAGCAGTGAATAGGCTTAGATAATATGATAAGGCTTTATGTA	1200
QY	1201	GTTGTGGAACAGGATTAGCAGATGCCATGAGTATAGCTTCAAGTTCCTTCTAATAATTA	1260
DB	1201	GTTGTGGAACAGGATTAGCAGATGCCATGAGTATAGCTTCAAGTTCCTTCTAATAATTA	1260
QY	1261	GATGTAATGTTGTTGTAGTAGAACAATGACATGCTTCTTCTAGGAAGGCTGATGTAGAT	1320
DB	1261	GATGTAATGTTGTTGTAGTAGAACAATGACATGCTTCTTCTAGGAAGGCTGATGTAGAT	1320
QY	1321	GGAAAGCTGATAAATATCTGATGACTTAGATAGTCTTCTAGGAAGGCTGATGTAGAT	1380
DB	1321	GGAAAGCTGATAAATATCTGATGACTTAGATAGTCTTCTAGGAAGGCTGATGTAGAT	1380
QY	1381	ATAATAGTGGATTTGCAAGTGTATCTGAAAGATGGAAGAGCTTATATCAGATGCTACT	1440
DB	1381	ATAATAGTGGATTTGCAAGTGTATCTGAAAGATGGAAGAGCTTATATCAGATGCTACT	1440
QY	1441	GTTAAAGCGTTTACAAAGATTTAAAGCGCAGATAGACACACACCTCTGAAGCTTATA	1500
DB	1441	GTTAAAGCGTTTACAAAGATTTAAAGCGCAGATAGACACACACCTCTGAAGCTTATA	1500
QY	1501	AAAAATATTATGCTTAATGATGATACTGAAATAGCTAAAGCTGCAAGTTTATAGATAAAGATTC	1560
DB	1501	AAAAATATTATGCTTAATGATGATACTGAAATAGCTAAAGCTGCAAGTTTATAGATAAAGATTC	1560
QY	1561	GGTCTTCAAGTAGTATGATGAGGAGTATTTAAATTTCTATGATGCTTAAAGATGATCTACA	1620
DB	1561	GGTCTTCAAGTAGTATGATGAGGAGTATTTAAATTTCTATGATGCTTAAAGATGATCTACA	1620
QY	1621	AAAGAGATCAATTAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1680
DB	1621	AAAGAGATCAATTAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1680
QY	1681	CCAGTTGTTATGCTTACTGTTTCTTCTGATCAATCGTTGCTGATGATGATGATGATGATGAT	1740
DB	1681	CCAGTTGTTATGCTTACTGTTTCTTCTGATCAATCGTTGCTGATGATGATGATGATGATGAT	1740
QY	1741	GTAGAGAAAAATATTCTAAAGATTTAAACAGATTTGCTCAAGGATAGCTTAATTCAGTT	1800
DB	1741	GTAGAGAAAAATATTCTAAAGATTTAAACAGATTTGCTCAAGGATAGCTTAATTCAGTT	1800
QY	1801	ATAAACAATAAGAAAGATTTTATGATATG 1830	
DB	1801	ATAAACAATAAGAAAGATTTTATGATATG 1830	

Db 1261 GATGGTAATGGTGTCTAGATAGAACAAATGGACATGCTACTCCAAATAGTTCTGTCTAGAT 1320
 QY 1321 GGAAAAGCTGATAAAATATCTGATGACTTAGATAGTTCTTAGGAAGCGCTGATGTAGAT 1380
 Db 1321 GGAAAAGCTGATAAAATATCTGATGACTTAGATAGTTCTTAGGAAGCGCTGATGTAGAT 1380
 QY 1381 ATAATAGGTGGATTGTGCAAGTGTATCTGAAAAGATGGAAGAGCTATATCAGATGCTACT 1440
 Db 1381 ATAATAGGTGGATTGTGCAAGTGTATCTGAAAAGATGGAAGAGCTATATCAGATGCTACT 1440
 QY 1441 GGTAAAGGGCTTTACAGAGCTTAAAGCGACGATAGACAAAGACACTAACTCTGAACTTATA 1500
 Db 1441 GGTAAAGGGCTTTACAGAGCTTAAAGCGACGATAGACAAAGACACTAACTCTGAACTTATA 1500
 QY 1501 AAAACATATATGCTAAATGATGACTGAAATAGCTAAAGCTGCAAGTGTATAGATAAAGATTCA 1560
 Db 1501 AAAACATATATGCTAAATGATGACTGAAATAGCTAAAGCTGCAAGTGTATAGATAAAGATTCA 1560
 QY 1561 GGTGCTTCAAGTAGTAGTCAGAGGATTTAAATTTCTATGTAGCTAAAGATGATCTACA 1620
 Db 1561 GGTGCTTCAAGTAGTAGTCAGAGGATTTAAATTTCTATGTAGCTAAAGATGATCTACA 1620
 QY 1621 AAAGAAGATCAATTAGTTGATGCAATTAGCAGTAGGAGCTGTTGCTGATATAAACTTGTCT 1680
 Db 1621 AAAGAAGATCAATTAGTTGATGCAATTAGCAGTAGGAGCTGTTGCTGATATAAACTTGTCT 1680
 QY 1681 CAGTTGTATAGTACTGATTTCTTTATCTTCTGATCAATCGTGTATAGCAAAAGTT 1740
 Db 1681 CAGTTGTATAGTACTGATTTCTTTATCTTCTGATCAATCGTGTATAGCAAAAGTT 1740
 QY 1741 GTAGGAGAAAATATTTCTAAAGATTTAACACAAAGTTGGTCAAGGAATAGCTAAATCAGTT 1800
 Db 1741 GTAGGAGAAAATATTTCTAAAGATTTAACACAAAGTTGGTCAAGGAATAGCTAAATCAGTT 1800
 QY 1801 ATAAACAAAATGAAAGATTTATTAGATATG 1830
 Db 1801 ATAAACAAAATGAAAGATTTATTAGATATG 1830
 RESULT 3
 ABS5367
 ID ABS5367 standard; DNA; 2268 BP.
 XX
 AC ABS5367;
 XX
 DT 15-NOV-2002 (first entry)
 XX
 DE Clostridium difficile strain 171500 slpA gene.
 XX
 KW Vaccine; Clostridium difficile; immunogenic; immunoreactivity;
 KW C. difficile infection; immune response; C. difficile associated disease;
 KW passive immunotherapy; passive vaccination; interleukin; adjuvant;
 KW C. difficile vaccine; antibacterial; slpA; gene; ds.
 XX
 OS Clostridium difficile.
 XX
 PN WO200262379-A2.
 XX
 PD 15-AUG-2002.
 XX
 PF 11-FEB-2002; 2002WO-IE000017.
 XX
 PR 09-FEB-2001; 2001IE-00000137.
 XX
 XX (QUEB-) QUEEN ELIZABETH COLLEGE DUBLIN.
 XX
 XX Doyle R, Kelleher D, Windle HJ, Walsh JB, Deirdre NE;
 XX
 XX WPI; 2002-643389/69.
 DR P-PSDB; ABG80346.
 XX
 XX Novel vaccine useful for treatment/prophylaxis of Clostridium difficile
 PT associated diseases, comprises Clostridium difficile genes or

PT peptide/polypeptides or its derivative, fragment, mutant or variant.
 XX Claim 44; Page 25-28; 85pp; English.
 PS
 CC The present invention relates to a new vaccine for treatment/prophylaxis
 CC of Clostridium difficile associated disease. The vaccine of the invention
 CC comprises C. difficile gene or peptide/polypeptide or derivative,
 CC fragment, mutant or variant of the peptide/polypeptide which is
 CC immunogenic in humans or to which immunoreactivity is detected in
 CC individuals who have recovered from C. difficile infection. The molecules
 CC of the invention are used in prophylaxis or treatment of C. difficile
 CC associated disease, where the medicament is a vaccine. The invention can
 CC be used by administering the vaccine preparation to a host to raise an
 CC immune response. The invention is further useful in the preparation of a
 CC medicament for treatment or prophylaxis of C. difficile infection or C.
 CC difficile associated disease. The invention is also useful in passive
 CC immunotherapy for established C. difficile infection, for the eradication
 CC of C. difficile associated disease, or for passive vaccination of an
 CC individual with C. difficile infection. Interleukin is useful as an
 CC adjuvant in C. difficile vaccine. The present nucleic acid sequence
 CC encodes a Clostridium difficile slpA protein, as described in the
 CC invention
 XX
 SQ Sequence 2268 BP; 893 A; 263 C; 443 G; 669 T; 0 U; 0 Other;
 Query Match 29.6%; Score 542; DB 6; Length 2268;
 Best Local Similarity 59.4%; Pred. No. 4.4e-90;
 Matches 1117; Conservative 0; Mismatches 670; Indels 93; Gaps 8;
 QY 32 CAGCTGTTTACTGTAGTAGTGTCTGCTCCAGTCTTTTGGCAGCAGCTTCAGATGTAATAT 91
 Db 401 CAGCAGATGCTATTAATGCTGGACATCTTCAGCAGATGGTGTGTTACAAATACTCGAG 460
 QY 92 CACTACAAGATGGTACAAATGATAGTATACAGTATCAAACTCAAACTAGTGTAGCTTAG 151
 Db 461 CTGCTAGTGGTCTTACTGAGACAAATTCAGCAGGAACAAAACCTTGCAATGTCACTATT 520
 QY 152 TAAAGGATATTTTAGCAGCACAAAACCTTAAACAACAGGTGCAGTTATTTTGAACAAAGATA 211
 Db 521 TTGACACAGCATATACAGATTCATCTGAAATTCGCGGTGAAGATTACTATAAAGCAGATA 580
 QY 212 CAAAAGTTTACT---TTCTATGATGCAATGAGAAAGATTTCTTCAACTCCAACTGGAGATA 268
 Db 581 TGAATGATCTAAATTTGGTAAAGCAGCTGAGACACTTATTCAACTGGGCTTACATTGG 640
 QY 269 AAAAGTTTATTCAGAACAAACTTTTAACTACAGCTAATGGAATGAAGATTTATGTAAGA 328
 Db 641 AAGATGGTCTACAGAAAAAATTTGTTAAATTTAGGGACAGTGATATTATAGATATAACTA 700
 QY 329 CAACCTTTAAAA---AATTTAGATGCGAGGAATATGCTATTATAGATTTAACTTATAATA 385
 Db 701 AAGCTCTTAAACTTACTGTTCTTCTGGAAGTAAAGCAACTGTTAAGTTTCTGTAAGAAAA 760
 QY 386 ATGCTAAACCTGTTG-----AAATTTAAAGTAGTAGCAGCTAGCTAGAAA 427
 Db 761 CACCAGTGCAGTGTTTCAACAGTAAACAACAGCTTAGAATAATAATGCTAAAGAAAG 820
 QY 428 AAACAGTAGTTGTATCTTAGTGATGCGAAAAAATAGTGCAAAAAGATATAGCTGAAAAAATAG 487
 Db 821 AAACAATAGATATTGACGCTAGTTCTAGTAAAAACAGCACAGATTTAGCTAAAAAATATG 880
 QY 488 TGTTTGAAGACAAAGACTTTAGAAAATGCATTAATAAACTATATAAGTCCCTCAGATTTACATA 547
 Db 881 TATTTAATAAACTGATTTAAATATCTCTTTATAAGTATTAATAGGATGAAGCAGATA 940
 QY 548 AAACGTAGTTTACTTATCAAGTAGTTCTTTTATCAAAAGGAAAGAGATTACAAGGTTTCT 607
 Db 941 CTAATGGATTAAAGAGAAGTAGTGGAAAATATCAAGTAGTTCTTTTATCCAGAAGAA 1000
 QY 608 CAACCTTATAGAGCTACAAATTTAT---AATGAAGGAACCTGCATATGTTTAAATACACCACTAA 664
 Db 1001 AAAGAGTTTAACTAAGAGTGTGCAAGGCTTCAATTTGCTGTATGAAAAATTCACCAGTTA 1060

Db	2129	TAGCTACTGATTCCTTTATCTTTCAGATCAAGTGTATCTATATAAGTAAAGTTCTTGATAAAG	2189					
Qy	1751	AATATATTCTAAAGATTTAAACACAAGTTGGTCAAGGAATAGCTAATTCAGTTATATAACAAAA	1810					
Db	2189	ATAATCGAGAAAACTTAGTTTCAAGTTGGTAAAGGTATAGCTACTTTCAGTTATATAACAAAA	2248					
Qy	1811	TGAAGATTTATTAGATATG	1830					
Db	2249	TGAAGATTTATTAGATATG	2268					
RESULT 4								
ABS65378	ABS65378 standard; DNA; 2271 BP.							
AC	ABS65378;							
XX	15-NOV-2002	(first entry)						
DT	Clostridium difficile strain 171448 slpA DNA sequence.							
XX	Vaccine; Clostridium difficile; immunogenic; immunoreactivity;							
XX	C. difficile infection; immune response; C. difficile associated disease;							
KW	passive immunotherapy; passive vaccination; interleukin; adjuvant;							
KW	C. difficile vaccine; antibacterial; slpA; ds.							
XX	Clostridium difficile.							
OS	WO200262379-A2.							
XX	15-AUG-2002.							
FD	11-FEB-2002; 2002WO-IE000017.							
PF	09-FEB-2001; 2001IE-00000137.							
XX	(QUEEN) QUEEN ELIZABETH COLLEGE DUBLIN.							
FA	Doyle R, Kelleher D, Windle HJ, Walsh JB, Deirdre NE;							
XX	WPI; 2002-643389/69.							
DR	Novel vaccine useful for treatment/prophylaxis of Clostridium difficile							
PT	associated diseases, comprises Clostridium difficile genes or							
PT	peptide/polypeptides or its derivative, fragment, mutant or variant.							
XX	Claim 14; Page 73-74; 85pp; English.							
PS	The present invention relates to a new vaccine for treatment/prophylaxis							
XX	of Clostridium difficile associated disease. The vaccine of the invention							
CC	comprises C. difficile gene or peptide/polypeptide or derivative,							
CC	fragment, mutant or variant of the peptide/polypeptide which is							
CC	immunogenic in humans or to which immunoreactivity is detected in							
CC	individuals who have recovered from C. difficile infection. The molecules							
CC	of the invention are used in prophylaxis or treatment of C. difficile							
CC	associated disease, where the medicament is a vaccine. The invention can							
CC	be used by administering the vaccine preparation to a host to raise an							
CC	immune response. The invention is further useful in the preparation of a							
CC	medicament for treatment or prophylaxis of C. difficile infection or C.							
CC	difficile associated disease. The invention is also useful in passive							
CC	immunotherapy for established C. difficile infection, for the eradication							
CC	of C. difficile associated disease, or for passive vaccination of an							
CC	individual with C. difficile infection. Interleukin is useful as an							
CC	adjuvant in C. difficile vaccine. The present nucleic acid sequence							
CC	represents a Clostridium difficile slpA DNA sequence of the invention							
XX	Sequence 2271 BP; 895 A; 263 C; 443 G; 670 T; 0 U; 0 Other;							
SQ								
Query Match 29.6%; Score 542; DB 6; Length 2271;								
Best Local Similarity 59.4%; Pred. No. 4.4e-90;								
Matches 1117; Conservative 0; Mismatches 670; Indels 93; Gaps 8								

Query Match 29.6%; Score 542; DB 6; Length 2271;
Best Local Similarity 59.4%; Pred. No. 4.4e-90;
Matches 1117; Conservative 0; Mismatches 670; Indels 93; Gaps 8

Qy	32	CAGCTGTTACTGTAGTAGGTTCTCGTGTCTCAGGTTTTTGCAGCAGCTTTCAGATGTAAATAT	91
Db	401	CAGCAGATGCTATATTTGCTGGAAACATCTTCAGCAGATGGTGTGTTGTACAAATACTGGAG	460
Qy	92	CAC'TACAAGATGGTACAAATGATAAGTATACAGTATCAATATCTAAAGCTTAGTGACTTAG	151
Db	461	CTGCTAGTGGTCTTACTTGAGACAAATTCAGCAGGAACAAACTTTCGAATGTCTAGCTATTTT	520
Qy	152	TAAAGGATATTTTAGCAGCAGCAAAACTTTAAACACAGGTGCAGTTATTTTTCGAAACAAAGATA	211
Db	521	TTGACACAGCATATACAGATTCATCTGAACTCGGGTTAAGTTACTATATAAAGCAGATA	580
Qy	212	CAAAAGTTACT--TTCTATGATGCAATGAGAAAGATTCTTCAACTCCCACTGGAGATA	268
Db	581	TGAATGATACTAAATTTGGTAAAGCAGCTGAGACAACTTATTTCACTGGGCTTACATTTG	640
Qy	269	AAAAAGTTTATTCAGAACAAACTTTAACTACAGCTAATGGAATGGAAGATTATGTAAGA	328
Db	641	AAGATGGGCTTACAGAAAAAATTGTTAAATTTAGGGGACAGTGATATTTATAGATATACTA	700
Qy	329	CAACTTTAAAA--AA'TTTAGATCAGAGAAATATGCTATTTATAGATTTTAACTTTATAATA	385
Db	701	AAGCTCTTAACTTACTTGTGTTCTCGAAGTAAAGCAACTGTTAAGTTTGTCTGAAAAAAA	760
Qy	386	ATGCTAAAACTGTG-----AAATTAAGTAGTAGCAGCTAGTAGTGA	427
Db	761	CACCAAGTGCAGTGTTCACCCAGTAATAACAAAGCTTAGAATAATAATGCTAAAGAAG	820
Qy	428	AAACAGTAGTTGTATCTAGTGATCGGAAAAATAGTGCAAAAAGATATAGCTGAAAAATATG	487
Db	821	AAACAAATAGATATTTGACGCTAGTTCTTAGTAAAAACAGCAAGATTTTAGCTTAAAAAATATG	880
Qy	488	TGTTTGAAGACAAAGACTTGAAGAAATGCAC'TAAAAACTATAAATGCGCTCAGATTTTCAGTA	547
Db	881	TATTTAATAAAC'TGATTTAAATACTCTTTATAAAGTATTAAATGGAGATGAACGAGATA	940
Qy	548	AAACTGATAGTTACTATCAAGTAGTTC'TTTATCCAAAAGGAAAGAGATTACAAGGTTTCT	607
Db	941	CTAATGGATTAAATAGAAGAGTTAGTGGAAAAATATCAAGTAGTTCTTTATCCAGAAGGAA	1000
Qy	608	CAACTTATAGAGCTACAAATAT--AATGAGGAAC'TGCATATGTTAATPACACCGATA	664
Db	1001	AAAGAGTTACAAC'TAAGAGTGTCTCAAGGCTTCAATTTGCTGATGAAAAATTCACCAAGTTA	1060
Qy	665	TATTAAC'TTAAAACTACTAGTAGAAGAGTAA'TTTTAAAGACTGCAGTAGAAGAGTTACAA	724
Db	1061	AA'TTAACTCTTAAGTCAGATAAGAAAGAAAGACTTAAAGAAATATGTGGATGATTTAAGAA	1120
Qy	725	AA'TGAAATGCTAGTTATTTCTAATACTACAAC'TTTAGCTGGTGTATGACAGAAATCAACAG	784
Db	1121	CATATAAATGGATATTTCAAATGCTATAGAGTAGCAGGAGAAGATAGAAATAGMAACTG	1180
Qy	785	CTATAGAGATAAGTAAAGAAATATTPACAATATATGATGGCGAGAAATCAGATCATTCAGCTG	844
Db	1181	CAATAGCATTTAAGTCAAAAAATTTATTAAC'TCTGATGATGAAAA-----TGCTA	1228
Qy	845	ATGTTAAAGAGAA'TGTTAAAAATGCTTATTTAGTAGGTGCAGTCAATGCATAGTAGTGGAT	904
Db	1229	TATTTAGAGATTGAGTTGATTAATGTATGTTGGTGGAGAAATGCAATAGTTGATGGAC	1288
Qy	905	TAGTTGGGGCTCCTTTTAGCAGCAGAAAAAGATGTCACATATTTATTTAACTTCAAAAAGATA	964
Db	1289	TTGTAGCTTCTCCTTTAGCTTCTGAAAGAAAGCTCTTTTATTTAACTTCAAAGATA	1348
Qy	965	AATTAGATTCTGTAGTAAAAATCTGAAATAAAGAGAGTTTTTAGACTTTAAAACTTCAACAG	1024
Db	1349	AA'TTAGATTCAAGCGTAAAAAGCTGAAATAAAGAGAGTTATGCAATATAAAGAGTACAACAG	1408
Qy	1025	AAGTA--ACAGGAAAAACAGTTTATATAGCTGTGGAGTTAATAGTGTATCTTAAAGAAG	1081
Db	1409	GTATAAATACTTCAAAGAAAGTTTATTTTAGCTGTGGAGTTAATCTTATATCTTAAAGAAG	1468
Qy	1082	TTGTAAACAGAAATTAGAATCAATGGGA'TTTAAAGTTTGAAGAGATTCTCAGGTGATGATAGAT	1141

Db	1469	TAGAAAATGAAATTAAAAAGATATGGGACTTTAAAGTTACAAGATTAGCAGGAGATGATAGAT	1528
Qy	1142	ATGAAACTCTCTTTAAAAATAGCAGTGAATAGCTTTAGATAATGATAAGGCTTTATCTAG	1201
Db	1529	ATGAACTCTCTTTAAAAATAGCTGATGAAGTAGGTCTTGATAATGATAAAGCAATTTGTAG	1588
Qy	1202	TTGGTGGAAACAGGATTAGCAGATGCCATGAGTATAGCTTCAGTTGGCTTCTACTAAATTTAG	1261
Db	1589	TTGGAGGAAACAGGATTAGCAGATGCCATGAGTATAGCTCCAGTTGCATCTCAATTAAGAA	1648
Qy	1262	ATGGTAATGGTGTGTAGATAGAACAAATGGACATGCTACTCCAAATAGTTGTTGTAGATG	1321
Db	1649	ATGCTAATGGTAAAAATGGATTTAGCTGATGGTATAGCTACCAATAGTAGTTGTAGATG	1708
Qy	1322	GAAAGCTGATAAAATATCTGATGACCTAGATAGTTTCTTAGGAAGCGCTGATGATAGATA	1381
Db	1709	GAAAGCTTAAACTTATAAATGATGATGATAAAGATTTCTTAGATGATTCACAAGTTGATA	1768
Qy	1382	TAATAGGTGGATTTGCAAGTGTATCTGAAAGATGGAAGAAGCTATATCAGATGCTACTG	1441
Db	1769	TAATAGGTGGGAAAAACAGTGTATCTAAAGATGTTGAAATGCAATAGATGATGCTACAG	1828
Qy	1442	GTAAGGCGTTTAAAGAGTTAAAGGCGACGATAGACAAGACATACTCTGAAAGTTATAA	1501
Db	1829	GTAATCTCCAGATAGATATAGTGGAGATGATAGACAAGCAACTAATGCCAAAAGTTATAA	1888
Qy	1502	AAACATATTATGCTAATGATACTGMAATAGCTAAAGCTGCAGTTTTAGATAAAGATTCAG	1561
Db	1889	AAGAACTCTCTTATTATCAAGATAACTTAATAATGATAAAAAAGTAGTTAATTTCTTTG	1948
Qy	1562	GTGCTTCAAGTAGTGATGAGGAGTAGTATTTAAATTTCTATGTAGCTAAAGATGGATCTACAA	1621
Db	1949	TAGCTAAGATGGTCTCTACTAAAGAGATCAATTAGTTGATGCTTTAGCAGCAGCTCCAG	2008
Qy	1622	A-----AGAAAGTCMAATTAGTTG	1639
Db	2009	TTGCAGCAAACTTTGGTGTAACTTTAATCTGATGGTAAGCCAGTAGATAAAGATGGTA	2068
Qy	1640	ATGCATTAGCAGTAGGAGCTGTTGCTGGATATAA-----CTTGCTCCAGTTGTAT	1690
Db	2069	AAGTATTAACTGGTCTTGTAATGATAAATAAATTAAGTATCCAGCACCTTATAGTAT	2128
Qy	1691	TAGCTACTGATCTTTTATCTTCTGATCAATCGGTTGCTATAAGCAAAAGTTGTAGGAGAA	1750
Db	2129	TAGCTACTGATCTTTTATCTTCTGATCAAAAGTGTATCTATAGTAAGTCTTGTATTAAG	2188
Qy	1751	AATATCTTAAGATTTTAACAACAAGTTGGTCAAGGAATAGCTAATTCAGTTATATAACAAA	1810
Db	2189	ATAATGGAGAAAACTTAGTTCAAGTTGGTAAAGGTATAGCTACTTCCAGTTATAACAAA	2248
Qy	1811	TGAAAAGATTTATTAGATATG	1830
Db	2249	TGAAAAGATTTATTAGATATG	2268
RESULT 5			
AAI65841			
ID	AAI65841 standard; DNA; 2271 BP.		
XX			
AC	AAI65841;		
XX			
DT	03-JAN-2002 (first entry)		
XX			
DE	Nucleotide sequence of S-layer protein of <i>C. difficile</i> strain 1.		
XX			
KW	Surface layer protein; S-layer protein; pseudomembranous colitis; PMC;		
KW	cell wall protein; gastrointestinal illness; abscess; wound infection;		
KW	osteomyelitis; urogenital tract infection; septicemia; peritonitis;		
KW	pleuritis; BS.		
XX			
OS	Clostridium difficile.		
XX			

Nucleotide sequence of S-layer protein of *C. difficile* strain 1.

Surface layer protein; S-layer protein; pseudomembranous colitis; PMC; cell wall protein; gastrointestinal illness; abscess; wound infection; osteomyelitis; urogenital tract infection; septicemia; peritonitis; pleuritis; ss.

Clostridium difficile.

PH Key Location/Qualifiers
 FT CDS 1..2271
 FT /tag= a
 FT /product= "S-layer protein"
 XX WC200173040-A1.
 XX
 XX
 XX PD 04-OCT-2001.
 XX
 XX
 XX PF 23-MAR-2001; 2001WO-GB001305.
 XX
 XX PR 24-MAR-2000; 2000GB-00007263.
 XX
 XX PA (UNLO) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
 XX
 XX PI Pairweather NF, Calabi E;
 XX WPI; 2001-616508/71.
 XX P-PSDB; AAG79242.
 XX
 XX PT Novel polypeptides and polynucleotides of cell wall proteins of
 PT Clostridium difficile especially S-layer cell wall protein useful for
 PT preventing and treating the infection caused by the bacteria.
 XX
 XX PS Claim 3; Page 57-58; 62pp; English.
 XX
 XX CC The present sequence encodes a surface layer (S-layer) protein of
 CC Clostridium difficile. The S-layer proteins are the predominant cell wall
 CC protein. There are two distinct S-layer proteins in C. difficile, a 45
 CC kDa and 36 kDa protein. S-layer polypeptides and polynucleotides are
 CC useful for treating and/or preventing a disease associated with
 CC C. difficile infection in a subject. Such diseases include
 CC pseudomembranous colitis (PMC) in humans characterized by diarrhoea, a
 CC severe inflammation of the colonic mucosa, and formation of
 CC pseudomembranes that are composed of fibrin, mucus, necrotic epithelial
 CC cells and leukocytes; gastrointestinal illness, abscesses, wound
 CC infections, osteomyelitis, urogenital tract infections, septicemia,
 CC peritonitis, and pleuritis
 XX
 XX SQ Sequence 2271 BP; 895 A; 263 C; 442 G; 671 T; 0 U; 0 Other;
 Query Match 29.3%; Score 535.6; DB 4; Length 2271;
 Best Local Similarity 59.2%; Pred. No. 6.6e-89;
 Matches 1113; Conservative 0; Mismatches 674; Indels 93; Gaps 8;
 32 CAGCTGTTACTGTAGTAGTTCGCTCCAGTTTTTGCAGCAGCTTCAGATGTAATAT 91
 401 CAGCAGATGCTATAATTGCTGGAACATCTTCAGCAGATGGTGTGTTTACAAATACTGGAG 460
 92 CACTACAAGATGGTACAAATGATATACAGTATCAATATCAATATCAATATGACTTAG 151
 461 CTGCTAGTGGTTCTACTGAGACAAATTCAGCAGGACAAACTTGCATGTCTAGCTATTT 520
 152 TAAAGGATATTTTACGACGACAAACTTAAACAGCTGCAGTTATTTTGAACAAAGATA 211
 521 TTGACACAGCATATACAGATTCATCTGAACTGGGTTAAGTTACTATTAAGCAGATA 580
 212 CAAAGTTACT---TTCTATGATGCAATGAGAAAGATTTCTCAACTCCAACTGGAGATA 268
 581 TGAATGATCTAAATTTGGTAAAGCAGGTGAGACAACTTATTCACCTGGGCTTACATTTG 640
 269 AAAAGTTTATTGAGACAAACTTTTAACTACAGCTATGGAATGGAAGATTTGTAAGA 328
 641 AAGATGGGTCTACAGAAAAAATTTGTTAAATTTAGGGGACAGTGATTTATATAGATATACTA 700
 329 CAACCTTTAAAR---AATTTAGATCCAGAGATATGCTATTATAGATTTAACTTATAATA 385
 701 AAGCTCTTAACTTACTGTGTTCTGGAAGTAAGCAACTGTTTAAAGTTTCTGCAAAAAA 760
 386 ATGCTAAAACCTGTTG-----AAATTTAAAGTAGTAGCAGCTAGTCAAA 427
 761 CACCAAGTGCCAGTGTTCACACCAGTAATAACAAAGCTTAGAATAATAAATGCTAAAGAG 820

QY 428 AAACAGTAGTTCGTATCTAGTGTGCGAAATAATAGTGCAAAAGATATATAGCTGAAAATATG 487
 DB 821 AAACAATAGATATTGACGCTAGTCTTAGTAAAAACAGCAAGATTTAGCTTAAAAATATG 880
 QY 488 TGTTTGAGACAAAGACTTTAGAAATGCACTAAAACATATAAATGCCTCAGATTTTCAGTA 547
 DB 881 TATTTTAAATAAACTGATTTTAAATACTCTTTTATAAAGTATTAAATGGAGATGAACAGATA 940
 QY 548 AAACCTGATAGTTACTATCAAGTAGTTCCTTTATCCAAAAGGAAAGAGATTACAAGGTTTCT 607
 DB 941 CTAAATGGATTAAATAGAGAGATTAGTGGAAATATCAAGTAGTCTTTTATCCAGAGGAA 1000
 QY 608 CAACCTTATAGACTCAAAATTAAT---AATGAAGGAACTGCATATGTAATATACACAGTAA 664
 DB 1001 AAAGAGTTACAACTAAGAGTGTGCAAGGCTTCAATTCGTATGATGAAATTCACCAAGTTA 1060
 QY 665 TATTAACCTTAAATCTACTAGTAAAGAGTAATTTAAAGACTGCAGTAGAAGAGTTACAAA 724
 DB 1061 AATTAACTCTTAAGTCAGATAAAGAGAAAGACTTTAAAGATTATGTGGATGATTTAAGAA 1120
 QY 725 AATTGAATGCTAGTTATTCTTAATCTCACTCACTTTAGCTGGTGTATGACAGATACAAAACAG 784
 DB 1121 CATATATTAATGGATTTCAATGCTATAGAAGTAGCAGAGAGATAGATAAGAACTG 1180
 QY 785 CTATAGAGATAAGTAAAGAAATATTACAATAATGATGGCAGAGAAATCAGATTCATTCAGCTG 844
 DB 1181 CAATAGCATTAAGTCAAAAATATATAAATCTGATGATGAAA-----TGCTA 1228
 QY 845 ATGTTAAAGAGATGTTAAAAATGTTTATTTAGTAGTGCCTCAATGCACTAGTAGATGAT 904
 DB 1229 TATTTAGAGATTCAGTTGATATGTTAGTATGTTGGAGAAATGCAATGTTGATGGAC 1288
 QY 905 TAGTTGCGCTCTCTTTAGCAGCAGAAAAAGATGCTCCACTATTATTAACTTCAAAAGATA 964
 DB 1289 TTGTAGCTTCTCTTTAGCTTCTGAAAGAAAGCTCTTTTATTATTAACTTCAAAAGATA 1348
 QY 965 AATTAGATTCGTCAGTAAATCTGAAATAAAGAGAGTTTGTAGCTTAAAACTTCAACAG 1024
 DB 1349 AATTAGATTCAGCGTAAAGCTGAAATAAAGAGAGTTTGAATATAAAGAGTACAAACAG 1408
 QY 1025 AAGTA---ACAGGAAAAACAGTTTATATAGCTGGTGGAGTTAATAGTATCTATAAGAAG 1081
 DB 1409 GTATAAATATCTTCAAAAGAAAGTTTATTAGCTGGTGGAGTTAATCTTATATCTAAAGAAG 1468
 QY 1082 TTGTAACAGAAATTTAGAAATCAATGGGATTAAGCTTGAAGAGATTCCTCAGGTGATGATAGAT 1141
 DB 1469 TAGAAATGAATTTAAAGATATGCGGACTTAAAGTTACAGATTTAGCAGGAGATGATAGAT 1528
 QY 1142 ATGAAACTCTCTTTTAAATAATAGCAGGTGAAATAGGCTTTAGATAATGATAAGGCTTATGTAG 1201
 DB 1529 ATGAAACTCTCTTAAATAATAGCTGATGAGTAGGCTCTTGATAATGATAAAGCATTTGTAG 1588
 QY 1202 TTGGTGAACAGAGATTTAGCAGATGCCATGATATAGCTTTCAGTTGCTTCTACTATAAATAG 1261
 DB 1589 TTGAGGAAACAGGATTTAGCAGATGCCATGATATAGCTCCAGTTGCACTCAATTAAGAA 1648
 QY 1262 ATGCTAATGCTGTTGTAGATAGAAACAAATGGACATGCTACTCCAATAGTTGTTGTAGATG 1321
 DB 1649 ATGCTAATGCTAATAATGATTTAGCTGATGGTGTGCTACACCAATAGTATGTTGTAGATG 1708
 QY 1322 GAAAGCTGATAAATAATCTGATGACTTATAGTATGTTTCTTGAAGAGCGCTGATGTAGATA 1381
 DB 1709 GAAAGCTAATACTATAAATGATGATGATAAAGATTTCTTTAGATGATTTCAAGTTGATA 1768
 QY 1382 TAATAGGTGATTTGCAAGTGTATCTGAAAGATGGAAGAGCTATATATCAGATGCTACTG 1441
 DB 1769 TAATAGGTGAGAAAAACAGTGTATCTAAAGATGTTTGAAGATGCAATAGATGATGCTACAG 1828
 QY 1442 GTAAGCGGTTTACAGAGTTTAAAGGCGCAGATAGACAAAGACACTTAACTCTGAAAGTTATA 1501
 DB 1829 GTAAATCTCCAGATAGATATAGTGGAGATGATAGCAAGCAACTTAATGCAAAAGTTATA 1888
 QY 1502 AAACATATTATGCTAATGATACTGAAATAGCTTAAAGCTGCAGTTTGTAGATAAAGATTCAG 1561

DB 1889 AGAATCTTCTTATTATCAGATAACTTAATAATGATAAATAAGTAGTTATTTCTTTG 1948
QY 1562 GTGCTTCAAGTAGTAGCAGAGATTTTAAATTTCTATGAGCTAAAGATGATCTACAA 1621
DB 1949 TAGCTAAGAGATGGTCTTACTAAGAAGATCAATTAGTTGATGCTTTAGCAGCAGCTCCAG 2008
QY 1622 A-----
DB 2009 TTGAGCAAACTTTGGTGTAACTTTAAATCTGTGANGGTAAGCCAGTAGATAAGATGGTA 2068
QY 1640 ATGCATTAGCAGTAGGAGCTGTGCTGGATATAA-----CTTGCTCCAGTTGTAT 1690
DB 2069 AGATTAATACTGGTCTGATATGATAAATAATAATAGTATCTCAGCAGCTATAGTAT 2128
QY 1691 TAGCTACTGATCTTTATCTTCTGATCAATCGGTGCTATAAGCAAAAGTTGTAGGAGAA 1750
DB 2129 TAGCTACTGATCTTTATCTTCTGATCAATCAAGTGATCTATAAGTAAAGTTCTTGATAAG 2188
QY 1751 AATATTTCTAAGATTTTAAACAGTTGGTCAAGGATAGCTAATTCAGTTATATAACAAA 1810
DB 2189 ATAATGGAGAAAACCTTAGTTCAAGTTGGTAAAGGTATAGCTACTTTCAGTTATAAACAAT 2248
QY 1811 TGAAGATTTATTAGATATG 1830
DB 2249 TAAAGATTTATTAGTATG 2268

RESULT 6
ABS65372
ID ABS65372 standard; DNA; 2217 BP.
XX AC ABS65372;
XX DT 15-NOV-2002 (first entry)
XX DE Clostridium difficile strain 173644 slpA gene.
XX KW Vaccine; Clostridium difficile; immunogenic; immunoreactivity;
KW C. difficile infection; immune response; C. difficile associated disease;
KW passive immunotherapy; passive vaccination; interleukin; adjuvant;
KW C. difficile vaccine; antibacterial; slpA; gene; ds.
XX OS Clostridium difficile.
XX PN WO200262379-A2.
XX PD 15-AUG-2002.
XX PF 11-FEB-2002; 2002WO-IE000017.
XX PR 09-FEB-2001; 2001IE-00000137.
XX PA (QUEE-) QUEEN ELIZABETH COLLEGE DUBLIN.
XX PI Doyle R, Kelleher D, Windle HJ, Walsh JB, Deirdre NE;
XX WPI; 2002-643389/69.
XX DR P-PSDB; ABG80351.
XX PT Novel vaccine useful for treatment/prophylaxis of Clostridium difficile
PT associated diseases, comprises Clostridium difficile genes or
PT peptide/polypeptides or its derivative, fragment, mutant or variant.
XX PS Claim 49; Page 45-48; 85pp; English.
XX CC The present invention relates to a new vaccine for treatment/prophylaxis
CC of Clostridium difficile associated disease. The vaccine of the invention
CC comprises C. difficile gene or peptide/polypeptide or derivative,
CC fragment, mutant or variant of the peptide/polypeptide which is
CC immunogenic in humans or to which immunoreactivity is detected in
CC individuals who have recovered from C. difficile infection. The molecules
CC of the invention are used in prophylaxis or treatment of C. difficile

CC associated disease, where the medicament is a vaccine. The invention can
CC be used by administering the vaccine preparation to a host to raise an
CC immune response. The invention is further useful in the preparation of a
CC medicament for treatment or prophylaxis of C. difficile infection or C.
CC difficile associated disease. The invention is also useful in passive
CC immunotherapy for established C. difficile infection, for the eradication
CC of C. difficile associated disease, or for passive vaccination of an
CC individual with C. difficile infection. Interleukin is useful as an
CC adjuvant in C. difficile vaccine. The present nucleic acid sequence
CC encodes a Clostridium difficile slpA protein, as described in the
CC invention
XX SQ Sequence 2217 BP; 856 A; 271 C; 432 G; 658 T; 0 U; 0 Other;
Query Match 27.5%; Score 503.4; DB 6; Length 2217;
Best Local Similarity 59.6%; Pred. No. 5.3e-83;
Matches 950; Conservative 0; Mismatches 616; Indels 27; Gaps 5;
QY 250 TCAACTCCAACTGGAGATAAAGAACTTTATTCAGAACAACTTTAACTACAGCTAATGGA 309
DB 640 TCTGTTCTCTACTACAGGCTTAACATTAACCTGCTGATACAACTGCAACACAGATGTAAT 699
QY 310 AATGAAGATTATGTAAAGACAACTTTAAAAAATTTAGATGCGAGGAGAATATCTATTATA 369
DB 700 ATTTCTGATGTTATGAGTGCAATTTAAATTTAATGCTACTGATACGATTAGTGGATCCCA 759
QY 370 GATTTAACTTATAAATGCTTAAACTGTTGAAATTTAAAGTAGTAGAGCTAGTGAGAAA 429
DB 760 GCTGGTTCATCAGCTTCTACTCTTAGAGCAAGTATAAAGTAATAAATGCAAAAGAGAA 819
QY 430 ACAGTAGTTGTTATCTAGTGATCGGAAAATAGTGCAAAAGATATAGCTGAAAAAATATG 489
DB 820 TCTATAGATTTGATTCAAGTTTACATAGAACAGCTGAAGATTTAGCTGAAAAAATATGTA 879
QY 490 TTTGAAGACAAAGACTTAGAAAAATGCACATAAAAAATATAAAATGCCTCAGATTTTC----- 543
DB 880 TTTAAACCAAGAGATGTGAATAAAACTTATGAGGCACTGACTGATTTATATAAAGAGGT 939
QY 544 AGTAAAACTGATGTTATCTATCAAGTAGTTCTTTTATCCAAAAGGAAGATTAACAAGGT 603
DB 940 ATAAACAAGTAATCTTATCACTCAAGATGGTGGAAAATATCAAGTTGTTTATTGCTCAA 999
QY 604 TTCTCAACTATAGAGCTACAAATTTAATGAAGNACTGTCATATGGTAATACACAGTA 663
DB 1000 GGAAGAGATTAACTACTATAAGGAGCAACTGGAACTTTAGCAGATGAAAAATTTCTCTCT 1059
QY 664 ATATTAACTCTAAAAATCTACTAGTAGAGTAAATTTAAAGACTGCGAGTAGAAGATTACAA 723
DB 1060 AAGTAAACAATAAAGCAGATAAAGTAAAGACTTTAAAGATTATGTTGAAGATTAAAA 1119
QY 724 AAATTTGAATGCTAGTTATTTCTAATACTACAACTTTAGCTGGTGATGACAGAAATACAA 783
DB 1120 AATGCTAAACAATGGATATTCAAAATTTCTGTTGTAGCAGGTGAAGATAGATAAGAA 1179
QY 784 GCTATAGAGATAAGTAAGAAATTTACAAATATGATGGCGGAATCAGATCATTCAGCT 843
DB 1180 GCAATAGAGTTAAGTAGCAAAATACATATACTCTCTGATGATGACAA-----TGCA 1227
QY 844 GATGTTAAAGAGAAATGTTAAAAAATCTGTATTAGTAGGTGCAAAATGCACTAGTAGATGA 903
DB 1228 ATAATTAAGATCCAGTTAAACAAATGTTTGTATTGTTGTTCTCAAGCTGTAGTTGATGG 1287
QY 904 TTAGTTGGGGTCCCTTTTAGCAGAGAAAAGATGTCCTCACTATATTATTAACCTTCAAAGAT 963
DB 1288 CTTGTAGCTTCACCTTTAGCATCTGAAAAAAGAGCTCTCTTTACTATTAACTTACAGCAGGA 1347
QY 964 AAATTTAGATTCCTCAGTAAATCTGAAATAAGAGAGTTTTAGACTTAAAACTTCAACA 1023
DB 1348 AAATTTAGATTTCAAGTTTAAAGCTGAGTTGAAAAGAGTAATGGATTTAAAAATCTCAACA 1407
QY 1024 GAAGTA---ACAGGAAAAACAGTTTATATAGCTGGTGGAGTTTAAATAGTGTATCTAAAGAA 1080
DB 1408 GGTGTAAATACTTCTTAAAAAAGTTTACTTTAGCTGGTGGAGTAAACTCTATATCTAAAGAT 1467

QY 724 AAATTGAATGCTAGTATTCTTAATCTACCAACTTTAGCTGGTGATGACAGAAATCAAAACA 783
 DB 1120 AATGCTAAACAATGGGATATTCAAAATCTGTTGTAGCAGGTTGAAGATAGATAGAAACA 1179
 QY 784 GCTATAGAGATAAGTAAGAAATATTACAAATATGATGGCGAGAAATCAGATCATTCAGCT 843
 DB 1180 GCAATAGATTAAAGTAGCAAAATCTATAAATCTGATGATGACAA-----TGCA 1227
 QY 844 GATGTTAAAGAGAATGTTAAAAATGTTGTTATAGTGGTGCAAAATGCACTAGTAGATGGA 903
 DB 1228 ATAACCTAAAGATCCAGTTAAACAATGTTGTTTATGTTGGTTCTCAAGCTGATGATGG 1287
 QY 904 TTAGTTGGGGCTCTTTAGCAGCAGAAAGAAAGATGCTCCACTATTATTAATCTTCAAAAGAT 963
 DB 1288 CTTGTAGCTTCACTTTAGTACTGCAAAAAGAGCTCTCTTTACTATTAACTTTACAGCAGA 1347
 QY 964 AAATTAGATTGCTCAGTAAATCTGAAATATAAGAGAGTTTATAGACTTAAAACTTCAACA 1023
 DB 1348 AAATTAGATTCAAGTGTAAAGCTGAGTTGAAAGAGTAATGGATTTAAATCTCAACA 1407
 QY 1024 GAAGTA---ACAGGAAAAACAGTTTATATAGCTGGTGGAGTTAATAGTGTATCTAAAGAA 1080
 DB 1408 GGTGTAATACTTCTAAAGAAAGTTTACTTAGCTGGTGGAGTAACTCTATACTTAAGAT 1467
 QY 1081 GTTGTAAACAGATTAGATCAATGGGATTTAAAGTTGAAAGATTTCTCAGGTGATGATAGA 1140
 DB 1468 GTAGAAAATGAATTAAGATATGGGACTTAAAGTTTACAAGATTATCAGGAGATGATAGA 1527
 QY 1141 TATGAACTTCTTTTAAAAATACAGCTGAAATAGGCTTAGATAAATGATAAGGCTTATGTA 1200
 DB 1528 TATGAACTTCTTTAGCTATAGCTGATGAAATAGGCTTGTGATTAATGATTAAGCTTTGTA 1587
 QY 1201 GTTGTGGAAACAGGATTAGCAGATGCTAGTATAGCTTTCAGTTGCTTCTCTAAATTA 1260
 DB 1588 GTTGGAGGAACAGGATTAGCGGATGCTATGATGATAGCTCCAGTTGCTTCTCAATTAAGA 1647
 QY 1261 GATGTAATGTTGTTAGATAGAACAAATGACATGCTACTCCAAATAGTTGTTGATAGAT 1320
 DB 1648 AACTCAAAATGGAGAACTTGA---CTTAAAGGTTGATGCAACTCCAAATAGTAGTTGAT 1704
 QY 1321 GGAAGAGCTGATAAATAATCTGATGACTTAGATAGTTTCTTAGGAAGCGCTGATGATAG 1380
 DB 1705 GGAAGAGCTGAAGATATAAATCTGAAGTAAAGATTTCTTAGATGATTCACAAAGTTGAT 1764
 QY 1381 ATAATAGTGGATTTTGCAGAGTGTATCTGAAAAGATGGAAGAGCTATATCAGATGCTACT 1440
 DB 1765 ATAATAGTGGTGTAAATAGTGTCTTAAAGAGTAATGGAAGCAATAGATGATGCTACT 1824
 QY 1441 GGTAAAGCGGTTTACAAGAGTTTAAAGCGACGATAGACAAAGACACTAACTCTGAAGTTATA 1500
 DB 1825 GGAATAATCACTCAGAGATATAGTGGAAGATAGACAAAGCAAAATGCTTAAAGTTATA 1884
 QY 1501 AAAACATATTATGCTAATGATCTGAAATAGCTTAAAGCTGCAAGTTTATAGATAAGATTCA 1560
 DB 1885 AAAGAGATGATTTCTTTAAATGGAAGATTTACAACTTCTTTGATGATTAAGATGTT 1944
 QY 1561 GGTGCTTCAAGTAGTATGATGATGATGATTTTAAATTTCTATGATGATTAAGATGATGAT 1620
 DB 1945 TCAACTTAAAGAGATCAATTAGTAGATGCTTTTAGCAGGTTGCTGCAATTCGCTGTAACTTT 2004
 QY 1621 AAAGAGATCAATTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
 DB 2005 GGTGTAACAGTATGATTAATGAAGGAAACCTACAGTTGCTGATTAAGGAAAGCTTCTCAGCA 2064
 QY 1681 CCAATTTGTTATGCTACTGATTTCTTTATCTTCTGATCAATCGTTGCTTATAGCAAGATTT 1740
 DB 2065 CCAATTTGTTATGCAACAGATTTCTTTATCTTCTGATCAATCGTTGCTTATAGCAAGATTT 2124
 QY 1741 GTA---GGAGAAAAATTTCTAAAGATTTTAAACACAAAGTTGGTCAAGGAATAGCTAATCA 1797
 DB 2125 GTAAATGATGACGCTAATCTAAGATCTAGTTTCAAGTTGGTAAAGGATAGCTACTTCA 2184

QY 1798 GTTATACAAAAATCAAGATTTATTATAGATATG 1830
 DB 2185 GTTGTAGTAAATAAAGATTTATTAGATATG 2217
 RESULT 8
 AAI65839
 ID AAI65839 standard; DNA; 2145 BP.
 XX AAI65839;
 AC AC
 DT 03-JAN-2002 (first entry)
 XX
 DE Nucleotide sequence of S-layer protein of *C. difficile* strain 17.
 XX
 KW Surface layer protein; S-layer protein; pseudomembranous colitis; PMC;
 KW cell wall protein; gastrointestinal illness; abscess; wound infection;
 KW osteomyelitis; urogenital tract infection; septicemia; peritonitis;
 KW pleuritis; ss.
 XX
 OS Clostridium difficile.
 XX
 FH Key Location/Qualifiers
 CDS 1..2145
 FT /*tag= a
 FT /product= "S-layer protein"
 FT
 XX WO200173040-A1.
 PD 04-OCT-2001.
 XX
 PF 23-MAR-2001; 2001WO-GB0001305.
 XX
 PR 24-MAR-2000; 2000GB-00007263.
 XX
 PA (UNLO) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
 XX
 PI Fairweather NP, Calabi E;
 XX
 DR WPI; 2001-616508/71.
 DR P-PSDB; AAG79240.
 XX
 PT Novel polypeptides and polynucleotides of cell wall proteins of
 PT Clostridium difficile especially S-layer cell wall protein useful for
 PT preventing and treating the infection caused by the bacteria.
 XX
 PS Claim 3; Page 56-57; 62pp; English.
 XX
 CC The present sequence encodes a surface layer (S-layer) protein of
 CC Clostridium difficile. The S-layer proteins are the predominant cell wall
 CC protein. There are two distinct S-layer proteins in *C. difficile*, a 45
 CC kDa and 36 kDa protein. S-layer polypeptides and polynucleotides are
 CC useful for treating and/or preventing a disease associated with
 CC *C. difficile* infection in a subject. Such diseases include
 CC pseudomembranous colitis (PMC) in humans characterized by diarrhoea, a
 CC severe inflammation of the colonic mucosa, and formation of
 CC pseudomembranes that are composed of fibrin, mucus, necrotic epithelial
 CC cells and leukocytes; gastrointestinal illness, abscesses, wound
 CC infections, osteomyelitis, urogenital tract infections, septicemia,
 CC peritonitis, and pleuritis
 XX
 SQ Sequence 2145 BP; 846 A; 261 C; 422 G; 616 T; 0 U; 0 Other;
 Query Match 26.4%; Score 483.2; DB 4; Length 2145;
 Best Local Similarity 62.7%; Pred. NO. 2.7e-79;
 Matches 911; Conservative 0; Mismatches 433; Indels 108; Gaps 6;
 QY 394 ACTGTTGAATTAAGTAGTAGCAGCTAGTGAAGAAACAGTAGTTGTATCTAGTGTGCG 453
 DB 784 ACTATAAAGTTAGAGTTACAGTGCAGAAAGAAAGTAATCTATTGATGTTGGAATTCAGATTCA 843
 QY 454 AAAAATAGTGCAGAAAGATATAGCTGAAAAATATGTTTGAAGACAAAGACTTAGAAAT 513

844	DB	TATATTAGTGCTGAAAAATTTT	AGCTAA	AAATATGTA	TTTAA	TCTCTAAAGGAGTTTCTGAA	903	
514	QY	GCAC	TAAAAAC	TATAAATGCCTCAGAT	TTTCAGTAA	AAAC	TGATAGTTACTATCAAGTAGTT	573
904	DB	GC	TTATAATGCAATAGTTGC	ATTACAA	ATGATGGA	ATAGA	ATCTGATTTAGTACAATTA	963
574	QY	CTTTAT	CCAAAGGAAGAGATTACA	AGGTTTCAAGGTTTCTCA	ACTTATAG	---	AGCTACAAATTTAT	630
964	DB	GTTAA	ATGCGAAAAATATCA	AGGTTATTTCTAT	CCAGAAGG	AAAAAGAT	TAGAAACTAAATCT	1023
631	QY	AA	TGAAGGAAC	TGCATATGCTAATAC	CACAGTAATATTAA	CTCTAAAA	CTCTACTAGTAAG	690
1024	DB	GC	AGATATAATAGCTGAT	GCAGATAGTCC	AGCTAAAT	TAATACTATAA	AGCTAATAATTA	1083
691	QY	AG	TAAATTTAAAGAC	TGCAGTAGAAGAGTTACA	AAAAATTTGA	ATCTAGTTATTTCTA	ATACT	750
1084	DB	AA	AGATTTAA	AGATTTATGTAGATGATTT	AAAAACATACA	TAATACTTACT	CAAAATGTT	1143
751	QY	AC	AATTTAGCTGGTGATGAC	AGAAATACAAACAGCTAT	TAGAGATAGTAAGTA	ATAATTAC		810
1144	DB	GTA	CAGTAGCAGGAGAAGATAGA	ATAAGAACTGCTATAGA	ATTAAGTAGTAATAAT	TTAT		1203
811	QY	AA	TAAATGATCGCGAGAAATCAG	ATCAATTCAGCTGATGTT	AAAGAGAAATGTTAAAA	ATGTT		870
1204	DB	AA	TTCTGATGATATAAAATGCA	ATAACTGATGATGCGATTAAT	-----	AA	TATA	1251
871	QY	GT	ATTAGTAGTGCAAAATGCA	CTAGTAGATGGATTAGTTG	CGGCTCCTTTAG	CAGCAGAA		930
1252	DB	GT	ATTAGTTGGAATCTAC	ATCTATAGTTGATGGTCTT	GTTCATCACC	ATTAGCTTCAGAA		1311
931	QY	AA	AGATGCTCCACTATTATTA	CTTCAAAAGATAAATTAG	ATTCGTCAGTAA	AAATCTGAA		990
1312	DB	AAA	CAGCTCCATTTATTTAA	CTTCAAAAGATAAATTAG	ATTCATCAGTAA	AAATCTGAG		1371
991	QY	AT	AAAGAGAGCTTTTAGAC	TTTAAAACTTCAACAGAA	GTA--	ACAG	AAAAACAGTTTAT	1047
1372	DB	AT	AAAAAGAGTTATGA	ACTTAAAGATGATCTG	GTATAATAC	TCTCTAAAAAGTTTAT		1431
1048	QY	AT	AGCTGGTGGAGTTAATAG	TGTATCTATAAAGAAAGTTG	TAAACAGAAATTAG	AAATCAATGGGA		1107
1432	DB	TT	AGCTGGTGGAGTTAATCT	ATATCTATAAGATGTAG	AAAAATGAA	TTGAAAAATATGGGC		1491
1108	QY	TT	AAAGTTGAAAGATCTCAG	GTGATGATAGATATGAA	ACTTCTTTAAAA	AAATAGCAGGT		1167
1492	DB	CT	TAAAGTTACTAGATTTAT	CAGGAGAAGACAGATAC	GAAACTCTCTTTAG	CAATAGCTGAT		1551
1168	QY	GAA	ATAGGCTTAGATATGATAG	GTATAGGCTTATGTTG	TGCTGGAACAGATTTAG	CAGATGCC		1227
1552	DB	GAA	ATAGGCTTGTATATGAT	ATAAGCAATTTGATGTT	GGTGATCTGGAT	TAGCAGATGCT		1611
1228	QY	AT	GAGTATAGCTTCAGTTG	CTTCTACTATAAATTTAG	ATGTTAGTGTGTTG	TAGATAGAACA		1287
1612	DB	AT	GAGTATAGCTCAGTTG	CTCT-----	CAACTTAA			1644
1288	QY	AAT	GGACATGCTACTCCA	ATAGTTGTTGTAGATCG	AAAGCTGATATA	AAATATCTGATGAC		1347
1645	DB	GAT	GGAGATGCTACTCCA	ATAGTAGTTGTAGATG	GAAGCAAAAGATAA	AGTGAAT		1704
1348	QY	TT	AGATAGTTTCTTAGG	AGCGCTGATCTAGATATA	TAGTGATTTGCA	AGTGATCT		1407
1705	DB	GCT	TAAGAGTTTCTTAGG	AACTTCTGATGTTGATATA	TAGTGGAAAAAT	ATACGGTATCT		1764
1408	QY	GAA	AGAGATGGAAGAGCTAT	ATCAGATGCTACTCGT	ATAAGCGTTACA	AGAGTTTAAAGGC		1467
1765	DB	AA	AGATTTGAAGAGTCA	ATAGATAGTGC	AACTGGAAAACTCC	CAGATAGATAAGTGA		1824
1468	QY	GAC	GATAGACAAGACACT	TAAC	TCTGAAGTTATA	AAAAACATAT	TATGCTAATGATCTGAA	1527
1825	DB	GAT	GACACACAAGCAACT	ATATGCTGAA	TTTTTAAAA-----			1860
1528	QY	AT	AGCTAAAGCTGACAGTTT	TAGATAAAGATTCAGG	TGCTCTCAAGTAGTGAT	GACGAGTA		1587
1861	DB	-----	-----	-----	-----	-----	-----	1890

Qy	1588	TTTAAATTTCTATGCTAGCAATGATGGATCTTACAAAAGAAAGATCAATTAGTTGATGCATTA	1644
Db	1891	GTGAATTTACTTTTGTGCAAAAGATGGTTCTACTAAAGAAAGATCATTAGTAGATGCATTA	1950
Qy	1648	GCAGTAGGAGCTGTGTGCTGG-----ATATAAACTTTGCTCCAGCTGTGTATTAGCTACT	1698
Db	1951	GCAGCAGCAACCAATAGCAGGTAGATTTAAGGAGTCTCCAGCTCCAATCATCTAGCTACT	2010
Qy	1699	GATCTTTTATCTTCTGATCAATCGGTTTGCTATATAAGCAGAAAGTTGTAGGAGAAATATTTCT	1758
Db	2011	GATACCTTTATCTTCTGACCAAAATGTAGCTGTGAAGTAAAGCAGTTCCTTAAGATGGTGGGA	2070
Qy	1759	AAAGATTTAACACAAGTTTGGTCTCAAGGAATAGCTAAATTCAGTTTATAAACAAAAATGAAGAT	1818
Db	2071	ACTAACTTAGTTCAAGTAGGTAAAGGTATAGCTTCTTCAGTTTATAACAAAAATGAAGAT	2130
Qy	1819	TTATTAGATATG	1830
Db	2131	TTATTAGATATG	2142

RESULT 9

ABS65371
ID ABS65371 standard: DNA: 2145 BP.

AC ABS65371;

DT 15-NOV-2002 (first entry)

XX DE *Clostridium difficile* strain 171862 *slpA* gene.

XX vaccine: clostridium difficile; immunogenic; i

C. difficile infection; immune response; passive immunotherapy; passive vaccination; interleukin; adjuvant; KW
C. difficile infection; immune response; passive immunotherapy; interleukin; adjuvant; KW
C. difficile infection; immune response; passive immunotherapy; interleukin; adjuvant; KW
C. difficile infection; immune response; passive immunotherapy; interleukin; adjuvant; KW

XX OS Clostridium difficile.

XX PN WO200262379-A2.

15-AUG-2002

11-FEB-2002: 2002WO-IE0000017.

XX
PR 09-FEB-2001: 2001IE-000000137.

XX
PA (QUEER-) QUEEN ELIZABETH COLLEGE DUBLIN.

XX
PT Doyle R. Kelleher D. Windle HJ, Walsh JB, Deirdre NE;

XX
DR
WPT: 2002-643389/69.

DR P-PSDB; ABG80350.
XX

PT associated diseases, comprises Clostridium difficile genes or
 PT new-ide/polymetides or its derivative fragment, mutant or variant.
 PI NOVEL vaccine useful for treatment/prophylaxis of clostridiosis

XX
PS Claim 48: Page 41-44: 85pp: English:

The present invention relates to a new vaccine for treatment/prophylaxis of *Clostridium difficile* associated disease. The vaccine of the invention comprises *C. difficile* gene or peptide/polypeptide or derivative, fragment, mutant or variant of the peptide/polypeptide which is immunogenic in humans or to which immunoreactivity is detected in individuals who have recovered from *C. difficile* infection. The molecules of the invention are used in prophylaxis or treatment of *C. difficile* associated disease, where the medicament is a vaccine. The invention can be used by administering the vaccine preparation to a host to raise an immune response. The invention is further useful in the preparation of a medicament for treatment or prophylaxis of *C. difficile* infection or *C. difficile* associated disease. The invention is also useful in passive immunotherapy for established *C. difficile* infection, for the eradication

CC of *C. difficile* associated disease, or for passive vaccination of an
CC individual with *C. difficile* infection. Interleukin is useful as an
CC adjuvant in *C. difficile* vaccine. The present nucleic acid sequence
CC encodes a Clostridium difficile α protein, as described in the
CC invention

XX
SO Sequence 2145 BP: 844 A: 261 C: 424 G: 616 T: 0 U: 0 Other:

Query Match 26.2%; Score 480; DB 6; Length 2145;
Best Local Similarity 62.6%; Pred.No. 1e-78;
Matches 909; Conservative 0; Mismatches 435; Indels 10

Qy	394	ACTGTTGAAATTAAGTAGTAGCAGCTAGTAGGAAAAACAGTAGTTGTATCTTAGTGATGCG	453
Db	784	ACTATAAAAGTTAGAGTTACAAAGTGCAAAAGAACTCTATTGATGTGGATTCAAGTTCA	843
Qy	454	AAAATAAGTCAAAAGATATAGCTGAAAAATATCTGTTTGAAGAAGAAGACTTAGAAAAT	513
Db	844	TATATTAGTGTGAAAAATTTAGCTTAAAAAATATGTATTAAATCCTAAAGAGGTTTCTGAA	903
Qy	514	GCACATAAAACTATAATGGCTCAGATTTCAGTAAAACTGATAGTTACTATCAAGTAGTT	573
Db	904	GCTTATTAATGCAATAGTTGCATTACAAAATGATGGATAGAACTGATTAGTACAAATTA	963
Qy	574	CTTTATCCAAAAGAAAGAGATTACAAGTTTCTCAACTTATAG--AGCTACAAATTAT	630
Db	964	GTTAAATGGAAAAATCAAGTTATTTTCTATCCAGAAGGAAAAAGATTAGAAACTAAATCT	1023
Qy	631	AATGAAGGAACGCATATGTTATATACACCAAGTAAATATTAATCTTAAATCTCTACTAGTAAG	690
Db	1024	GCAGATATAATAGCTGATGAGCAGATAGTCCAGCTAAAAATTAATATAAAGCTAAATAATTA	1083
Qy	691	AGTAATTTAAAGACTGCGAGTAGAAGAGTTCAAAAAATTGAATGCTAGTATTCTTAATACT	750
Db	1084	AAAGATTTAAAGATTATGTAGATGATTTAAACATACAAATACTTACTCAATGTT	1143
Qy	751	ACAACTTTAGCTGGTGTATGACAGAAATACAAACAGCTATAGAGATAAGTAAAGAATATTAC	810
Db	1144	GTAACAGTAGCAGGAGAAGATAGAAATAGAACTGCTATAGAA'TTAAGTAGTAAATATTAT	1203
Qy	811	AATAATGATGGCGAANAATCAGATCATTCAGCTGATGTTTAAAGAGAATGTTAAAAATGTT	870
Db	1204	AA'TCTGATGATAAAAAATGCAATAACTGATGATGCAGTTAAT-----AATATA	1251
Qy	871	GTATTAGTAGTGCAAAATGCACTAGTAGATGGATTAGTTCGGCTCCTTTAGCAGCAGAA	930
Db	1252	GTATTAGTTGAATCTACATCTATAGTTGATGGTCTTGTTCATCACATTTAGCTTCAGAA	1311
Qy	931	AAAGATGCTCCACTATTATTAATCTCAAAAGATAAAATAGATTCGTGAGTAAAAATCTGAA	990
Db	1312	AAAAACGCTCCATTTATTATTAATCTTCAAAAGATAAAATAGATTCATCAGTAAAAATCTGAG	1371
Qy	991	ATAAAGAGAGTTTTAGACTTTAAAAAATTCAACAGAGTA---ACAGGAAAAACAGTTTAT	1047
Db	1372	ATAAAAAAGAGTTATGAATCTTAAAGAGTGAATCGTGTATAAATACTTCTAAAAAAGTTTAT	1431
Qy	1048	ATAGCTGCTGGAGTTAATAGTGTATCTTAAAGAGTCTGTAACAGAA'TTAGAATCAATGGGA	1107
Db	1432	TTAGCTGGTGGAGTTAATCTTATCTTAAAGATGTAGAGATGAATTTGAAAAATATGGGC	1491
Qy	1108	TTAAAAGTTGAAAGATTCTCAGGTGATGATAGATATGAAACTCTCTTTAAAAAATAGCAGGT	1167
Db	1492	CTTAAAGTTACTAGATTATCAGGAGAAGACAGATACGAACTCTTTTAGCAATAGCTGAT	1551
Qy	1168	GAAATAGGCTTAGATAATATGATAGGCTTATGTAGTTGGTGGAAACAGATTTAGCAGATGCC	1227
Db	1552	GAAATAGGCTCTTGATAATATGATAAAGCAATTTGTAGTTGGTGGTACTGCTGGCAGATGCT	1611
Qy	1228	ATGAGTATAGCTTCAGTTGCTTCTACTTAAATTAGATGCTTAATGGTGTGTGTAGATAGACA	1287
Db	1612	ATGAGTATAGCTCCAGTTGCTTCT-----CAACTTAAA	1644
Qy	1288	AATGGACATGCTACTCCCAATAGTTGTTGTAGTAGGAAAAAGCTGATAAAAATATCTGATGAC	1347

PT peptide/polypeptides or its derivative, fragment, mutant or variant.
 XX Claim 17; Page 77-78; 85pp; English.
 XX The present invention relates to a new vaccine for treatment/prophylaxis of Clostridium difficile associated disease. The vaccine of the invention comprises C. difficile gene or peptide/polypeptide or derivative, fragment, mutant or variant of the peptide/polypeptide which is immunogenic in humans or to which immunoreactivity is detected in individuals who have recovered from C. difficile infection. The molecules of the invention are used in prophylaxis or treatment of C. difficile associated disease, where the medicament is a vaccine. The invention can be used by administering the vaccine preparation to a host to raise an immune response. The invention is further useful in the preparation of a medicament for treatment or prophylaxis of C. difficile infection or C. difficile associated disease. The invention is also useful in passive immunotherapy for established C. difficile infection, for the eradication of C. difficile associated disease, or for passive vaccination of an individual with C. difficile infection. Interleukin is useful as an adjuvant in C. difficile vaccine. The present nucleic acid sequence represents a Clostridium difficile slpA DNA sequence of the invention

XX SQ Sequence 2145 BP; 844 A; 261 C; 424 G; 616 T; 0 U; 0 Other;

Query Match 26.2%; Score 480; DB 6; Length 2145;
 Best Local Similarity 62.6%; Pred. No. 1e-78;
 Matches 909; Conservative 0; Mismatches 435; Indels 108; Gaps 6;

QY 394 ACTGTTGAAATTAAGTAGTAGCAGCTAGTGAAGAAACAGTAGTCTATCTAGTATGCG 453
 DB 784 ACTATAAAGTTAGAGTTACAGTGCAGAAAGAGAAATCTATTGATGGATTCAAGTTCA 843

QY 454 AAAAATAGTGCAAAAGATATAGCTGAAATATATGTTTGAACACAAAGACTTAGAAAT 513
 DB 844 TATATTAGTCTGAAATTTAGCTAAATAATATATGTTTAAATCCTAAAGAGGTTCTGAA 903

QY 514 GCACATAAACTATAATGCTCAGATTTTCAGTAAACCTAGTATTACTATCAAGTAGTT 573
 DB 904 GCTTATAATGCAATAGTTGCAATTAACAATATGGAATAGATCTGATTTAGTACAAATTA 963

QY 574 CTTTATCCAAAGGAAGAGATTACAAGGTTTCTCAACTTATAG---AGTCAAAATTTAT 630
 DB 964 GTTAATGGAATATCAAGTTATTTCTATCCAGAGGAAAGAGATTAGAAACTAAATCT 1023

QY 631 AATGAAGAACTGCATATGTAATACACAGTAAATATTAATCTTAAATCTACTAGTAAG 690
 DB 1024 GCAGATATAATAGCTGATGAGAGATAGTCCAGCTAAATAACTATAAAGCTAAATA 1083

QY 691 AGTAATTTAAAGACTGCAGTAGAAGAGTTACAAAATTTGAATGCTAGTTATTTCTAAT 750
 DB 1084 AAGATTTAAAGATTATGATAGATGATTTAAACATACATTAATCTACTCAATGTT 1143

QY 751 ACACTTTAGCTGTGATGACAGAAATCAACAGCTATAGAGATAAGTAAGAAATTTAC 810
 DB 1144 GTAACAGTACAGGAGAGATAGATAGAACTGCTATAGAAATTAAGTAAATATTTAT 1203

QY 811 AATAATGATGGCGAAATCAGATTCATTCAGCTGATGTTTAAAGAGATGTTAAATGTT 870
 DB 1204 AATCTGATATAAAATGCAATAAATCTGATGATGATGATGATGATGATGATGATGAT 1251

QY 871 GTATTAGTGTGCAATGCACTAGTAGATGATGATGATGATGATGATGATGATGATGAT 930
 DB 1252 GTATTAGTGTGATCTACATCTATAGTTGATGATGATGATGATGATGATGATGATGAT 1311

QY 931 AAGATGCTCCACTATTATTAATCTTAAAGATAAATTAGATTGCTGAGTAAATCTGAA 990
 DB 1312 AAAACACTCCATTTATTTAACTTCAAGATAAATTAGATTGCTGAGTAAATCTGAG 1371

QY 991 ATAAGAGAGTTTATGACTTAAATCTTCAACAGAGTA---ACAGGAAACAGTTTAT 1047
 DB 1372 ATAAAGAGAGTTTATGAACTTAAAGAGTGATGATGATGATGATGATGATGATGATGAT 1431

QY 1048 ATAGCTGGTGGATTATAGTGTATCTAAAGAGTTGTTAACAAGATTAGAAATCAATGGGA 1107

DB 1432 TTAGCTGGTGGAGTTAAATTTCTATATCTAAAGATGTAGAAGATGAATTTGAAAATATGGGC 1491
 QY 1108 TTAAAGTTTGAAGATTTCTCAGGTGATGATAGATATGAACTTTCTTTTAAAAATAGCAGGT 1167
 DB 1492 CTTAAAGTTTACTAGATTATCAGGAGAGACAGATACGAACTTTCTTTAGCAATAGCTGAT 1551
 QY 1168 GAAATAGGCTTTAGATAAATGATAAGGCTTTATGTAGTTGGTGGAAACAGGATTAGCAGATGCC 1227
 DB 1552 GAAATAGGCTTTGATAAATGATAAAGCAATTTGTAGTTGGTGGTACTGGATTGGCAGATGCT 1611
 QY 1228 ATGAGTATAGCTTCAAGTTGCTTCTACTAAATTTAGATGGTAAATGGTCTTGTAGATAGAAC 1287
 DB 1612 ATGAGTATAGCTTCAAGTTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1644
 QY 1288 AATGACATGCTTACTCCTAATAGTTGTTGTAGATGGAAGCTGATAAATAATATCTGATGAC 1347
 DB 1645 GATGAGATGCTTACTCCTAATAGTTGTTGTAGATGGAAGCTGATAAATAATAGTATGAT 1704
 QY 1348 TTAGATAGTTTCTTAGGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1407
 DB 1705 GCTAAGAGTTTCTTAGGAACTTCTGATGTTGATATAATAGTGGAAAAAATAGCGTATCT 1764
 QY 1408 GAAAGATGGAAGAGCTTATATCAGATGCTACTGCTTAAAGCGTTACAGAGCTTAAAGGC 1467
 DB 1765 AAGAGATTGAAGAGTCAATAGATAGTGCACACTGGAAAAAATCCAGATAGAATAAGTGA 1824
 QY 1468 GAGGATAGACAGACACTAACTCTGAAGTTTATAAACAACATATTTATGCTAATGATGATGAA 1527
 DB 1825 GATGACAGACAGCAACTAATGCTGAAGTTTAAATCTCTCTCTCTCTCTCTCTCTCTCTCT 1860
 QY 1528 ATAGCTAAAGCTGCAGTTTATAGATAAGATTCAGGTGCTTCAAGTAGTATGATGATGATGAT 1587
 DB 1861 -----GAAGATGATTATTTCAAGAGATGCTGAAGTT 1890
 QY 1588 TTTAATTTCTATGCTTAAAGATGATCTACAAAAGAGATCAATTTAGTTGATGATGATGAT 1647
 DB 1891 GTGAATTTACTTTGTCAAAAGATGCTTCTTCTTAAAGAGATCAATTTAGTATGATGATGAT 1950
 QY 1648 CAGATGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1698
 DB 1951 GCAG 2010
 QY 1699 GATTTCTTTATCTTCTGATCAATGCTGCTTATAGCAAAAGTTGTAGAGAGAAAAATATCT 1758
 DB 2011 GATCTTTATCTTCTGACCAAAATGATGCTGTAAGTAAAGCAGTTCTCTAAAGATGGTGA 2070
 QY 1759 AAGATTTAACAAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1818
 DB 2071 ACTAATCTTAGTTCAAGTAGGTAAGGTATAGCTTCTTCTGATTTAATCAAAATGAAAGAT 2130
 QY 1819 TTATTAGATATG 1830
 DB 2131 TTATTAGATATG 2142

RESULT 11
 ABS65369 standard; DNA; 2157 BP.
 XX ABS65369;
 AC ABS65369;
 XX 15-NOV-2002 (first entry)
 DT Clostridium difficile strain 170324 slpA gene.
 XX Vaccine; Clostridium difficile; immunogenic; immunoreactivity;
 KW C. difficile infection; immune response; C. difficile associated disease;
 KW passive immunotherapy; passive vaccination; interleukin; adjuvant;
 KW C. difficile vaccine; antibacterial; slpA; gene; ds.
 XX Clostridium difficile.

OY

1784 GAATAGCTAAATTGTTTATAAACAAGAATGAAGATTTATTAGATATG 1830
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D6

2111 GTATAGCTCTTCAGTTATAAACAAGAATGAAGATTTATTAGATATG 2157

RESULT 12

ABS65374

ID ARS65374 standard: DNA: 2157 BP.

AC ABS65374:

15-NOV-2002 (first entry)

XX
DE *Clostridium difficile* strain 170426 *slpA* gene.

[illegible]

RV vaccine; Clostridium difficile; immunogenic; immunoreactivity; C difficile infection; immune response; C difficile associated disease;

interleukin: adjuvant: negative vaccination: negative immunotherapy: C. difficile infection; immune response; C: difficile associated disease

passive immunotherapy, passive vaccination, immunization

XX
XX
XX

XX
 DW
 W03003370-28

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

XX

XX
XX

XX

DR P-PSDB: ABG80353.

Novel vaccine useful for treatment/prophylaxis of Clostridium difficile associated diseases, comprises Clostridium difficile genes or polypeptides or its derivative fragment, mutant or variant.

XX
PC 01-04-05 51. Page 53-56: 8500. English

The present invention relates to a new vaccine for treatment/prophylaxis of Clostridium difficile associated disease. The vaccine of the invention comprises C. difficile gene or peptide/polypeptide or derivative, fragment, mutant or variant of the peptide/polypeptide which is immunogenic in humans or to which immunoreactivity is detected in individuals who have recovered from C. difficile infection. The molecules of the invention are used in prophylaxis or treatment of C. difficile associated disease, where the medicament is a vaccine. The invention can be used by administering the vaccine preparation to a host to raise an immune response. The invention is further useful in the preparation of a medicament for treatment or prophylaxis of C. difficile infection or C. difficile associated disease. The invention is also useful in passive immunotherapy for established C. difficile infection, for the eradication of C. difficile associated disease, or for passive vaccination of an individual with C. difficile infection. Interleukin is useful as an adjuvant in C. difficile vaccine. The present nucleic acid sequence encodes a Clostridium difficile sfpA protein, as described in the invention

sequence 2157 BP: 843 A: 244 C: 435 G: 635 T: 0 U: 0 Other;

Query Match 25.8% Score 471.8 DB 6 Length 2157;

Query Match	23:30, 2002
Best Local Similarity	61.5%
Pred. No.	3.3e-77

BEST LOCAL SIMILARITY 91.50, ID: 3036 7;
Matches 951: Conservative 0: Mismatches 467: Indels 129: Gaps 7;

296 CTACAGCTAATGGAAATGAAGATTATGTAAAGACAACTTTAAAAAATTAGATGCAGGAG 355

238 C A C C C T A C T C C T A G C T G T A A G T G G T T T G T A C T A A A G A T G A T A C T G A T T T A G C A A 787

CCTCCTCCGCGTGGTTTCATACCTGCTGCTGAAATTAAAGTAG 415

QY 836 ATTACGCTGATGTTAAAGAGAAATGTTAAATAATGTTGTTATTTAGTGTGCAAAATGCACCTAG 895
 Db 1232 AAAATGCAATACTGATAAGCGGTTAATGATATAGTATTAGTTGGATCTACATCTATAG 1291
 QY 896 TAGATGGAATAGTTGGCGCTCCCTTTAGACGACGAAAAAGATGCTCCACATATTATTAACCTT 955
 Db 1292 TTGATGGTCTGTTGTCATCACCATTAGCTTTCAGAAAAACAGCTCCATTTATTATTAACCTT 1351
 QY 956 CAAAGATATAATAGATTCGTGAGTAAATCTGAAATAAAGAGAGTTTATGAGCTTAAATAA 1015
 Db 1352 CAAAGATATAATAGATTCATCAGTAAATCTGAAATAAAGAGAGTTATGAACTTAAAGA 1411
 QY 1016 CTTCAACAGAGTA---ACAGGAAACAGATTATATATAGCTGCTGGAGTTAATAGTGTAT 1072
 Db 1412 GTGACACTGGTATAAATCTCTTAAAAAAAGTTTATTTAGCTGGTGGAGTTAATCTTATAT 1471
 QY 1073 CTAAGAAGTTGTAACAGAGATTAGAAATCAATGGGATTTAAAGTTTGAAGATTTCTCAGGTG 1132
 Db 1472 CTAAGAAGTTGTAAGAATGAATGAAACACATGGGCTTTAAAGTTACTAGATTATCAGGAG 1531
 QY 1133 ATGATAGATATAAAGCTTTTAAATAATAGCAGGTGAAATAGCTTTAGATATATGATAAGG 1192
 Db 1532 AAGACAGATACGAAACTTTCTTAGCAATAGCTGATGAAATAGCTCTTGATAATGATAAAG 1591
 QY 1193 CTTATGCTAGTTGGTGAACAGAGATTAGCAGATGCCATGATAGTATAGCTTCAGTTGCTTCTA 1252
 Db 1592 CATTTGATGTTGGTGTGCTGGAATAGCAGATGCTATGAGTATAGCTCCAGTTGCTTCT- 1650
 QY 1253 CTAATATTAGATGTAATGTTGTTAGATAGAGAACAAATGGACATGCTACTCCAATAGTTG 1312
 Db 1651 -----CAACTTAAAGATGAGATGCTACTCCAATAGTAG 1684
 QY 1313 TTGTAGATGGAAGCTGATATAAATATCTGATGACTTATGATAGTATGTTCTTAGGAAGCGCTG 1372
 Db 1685 TTGTAGATGGAAGCTGATATAAATATCTGATGACTTATGATAGTATGTTCTTAGGAAGCTG 1744
 QY 1373 ATGTAGATATAAGTGTGTTGCAAGTGTCTGAAAGATGGAAGAGTGAAGAGTCAATAGATA 1432
 Db 1745 ATGTAGATATAAGTGTGGAAGAAATAGCGTATCTTAAAGAGATTTGAAGAGTCAATAGATA 1804
 QY 1433 ATGCTACTGGTAAAGCGTTTCAAGAGTTTAAAGGCGACATAGACAAAGACACTAACTCTG 1492
 Db 1805 GTGCAACTGGAAGAACTCCAGATAGATAAGTGGAGATGATAGACAAAGCACTAATGCTG 1864
 QY 1493 AGTTTATAAACAATATATGCTTAATGATGATGACTGAAATAGCTAAAGCTGCAGTTTATAGATA 1552
 Db 1865 AAGTTTATAAAGAAAGATGATTATTTTC----- 1890
 QY 1553 AAGATTCAGGTGCTTCAAGTATGATGACGAGGATTTTAAATTTCTATGATGATTAAGATG 1612
 Db 1891 -----ACAGATGGTGAAGTTGTGTAATTTACTTTGTTGCAAAAGATG 1930
 QY 1613 GATCTACAAAGAAAGATCAATAGTTGATGCAATTAGCAGTAGGAGCTGTTGCTGG----- 1667
 Db 1931 GTTCTACTAAGAGATCAATAGTAGATGCTTACGACAGCAACCAATAGCAGGTAGAT 1990
 QY 1668 ----ATATAAATCTGCTCCAGTTGATATGATGCTACTGATTTCTTTATCTTCTGATCAATCGG 1723
 Db 1991 TTAAGGAGTCTCCAGCTCCAATCATACTAGCTACTGATGATCTTATCTTCTGACCAAAATG 2050
 QY 1724 TTGCTATAGCAAGTTGAGGAGAAATATCTTAAAGATTTAACAAGTTTGGTTCAG 1783
 Db 2051 TAGCTGTAAGTAAGCAGTTCTTAAAGATGGTGGAACTAATCTAGTTTCAAGTAGGTGAAG 2110
 QY 1784 GAATAGCTAATTCAGTTTATAAACAAGATTTTATTTATAGATG 1830
 Db 2111 GTATAGCTTCTTCAGTTATATAAACAAGATTTATAGATG 2157

RESULT 15
 ABS65377
 ID
 ABS65377 standard; DNA; 2158 BP.
 XX

AC ABS65377;
 XX 15-NOV-2002 (first entry)
 DT Clostridium difficile strain 170324 slpA DNA sequence.
 XX Vaccine; Clostridium difficile; immunogenic; immunoreactivity;
 XX C. difficile infection; immune response; C. difficile associated disease;
 KW passive immunotherapy; passive vaccination; interleukin; adjuvant;
 KW C. difficile vaccine; antibacterial; slpA; ds.
 XX Clostridium difficile.
 OS WO200262379-A2.
 XX 15-AUG-2002.
 PD 11-FEB-2002; 2002WO-IE000017.
 PF 09-FEB-2001; 2001IE-00000137.
 PR (QUBE-) QUEEN ELIZABETH COLLEGE DUBLIN.
 XX Doyle R, Kelleher D, Windle HJ, Walsh JB, Deirdre NE;
 PI WPI; 2002-643389/69.
 DR Novel vaccine useful for treatment/prophylaxis of Clostridium difficile
 PT associated diseases, comprises Clostridium difficile genes or
 PT peptide/polypeptides or its derivative, fragment, mutant or variant.
 XX Claim 13; Page 71-72; 85pp; English.

CC The present invention relates to a new vaccine for treatment/prophylaxis
 CC of Clostridium difficile associated disease. The vaccine of the invention
 CC comprises C. difficile gene or peptide/polypeptide or derivative,
 CC fragment, mutant or variant of the peptide/polypeptide which is
 CC immunogenic in humans or to which immunoreactivity is detected in
 CC individuals who have recovered from C. difficile infection. The molecules
 CC of the invention are used in prophylaxis or treatment of C. difficile
 CC associated disease, where the medicament is a vaccine. The invention can
 CC be used by administering the vaccine preparation to a host to raise an
 CC immune response. The invention is further useful in the preparation of a
 CC medicament for treatment or prophylaxis of C. difficile infection or C.
 CC difficile associated disease. The invention is also useful in passive
 CC immunotherapy for established C. difficile infection, for the eradication
 CC of C. difficile associated disease, or for passive vaccination of an
 CC individual with C. difficile infection. Interleukin is useful as an
 CC adjuvant in C. difficile vaccine. The present nucleic acid sequence
 CC represents a Clostridium difficile slpA DNA sequence of the invention
 XX Sequence 2158 BP; 843 A; 246 C; 435 G; 634 T; 0 U; 0 Other;
 SQ Query Match 25.8%; Score 471.8; DB 6; Length 2158;
 Best Local Similarity 61.5%; Pred No. 3.3e-77;
 Matches 951; Conservative 0; Mismatches 467; Indels 129; Gaps 7;
 QY 296 CTACAGCTAATGGAATGAAGATTATGTAAGACAACTTTAAAAATTTAGATGACGAG 355
 Db 728 CAACACCTAGTCTGCTAGCTGTAAGCTGTTTGTAACTAAAGATGATCTGATTTAGCAA 787
 QY 356 RATATGCTATTATAGATTTAACTTATATATGCTTAAACTTTGAAAATTAAGTAGTAG 415
 Db 788 AATCAGGTACTATAAATTAAGATTAAAGTTTAAATGCAAAAAGAAATCAATTTGATATAGATG 847
 QY 416 CAGCTAGTGAATAAAGACAGTAGTTGATCTAGTGAATGCGAAAAATAGTCAAAAGATATAG 475
 Db 848 CAGCTCATATACATCAGCTG-----AAATTTAGCTTAAAGATATGATTTTGTATCCAG 901
 QY 476 CTGAAAAATATGTTGTTTGAAGCAAAAGACTTTAGAAAAATGCCTAAAAACTATAAATGCCT 535
 Db 902 ATGAAATTTCTGAACATATAAGGCAATAGTACGATTACAAATGATGTTAGATGATCTA 961

QY 536 CAGATTTTCAGTAAACTGATAGTACTATCAAGTAGTCTTCTATCCAAAGGAAAGAGAT 595
 Db 962 ACTTAGTTCAGTTAGTTATGGAATAATCAAGTGATTTTATCCAGAGGTAAAGAT 1021
 QY 596 TACAGGTTTCTCAACTATATAGAGTACAAATATATATGAAGAACTGCATATGGTAATA 655
 Db 1022 TAGAACTAAATCAGCA-----AATGATACAATAGCTAGTCAAGATA 1063
 QY 656 CACAGTATATTAATCTTAATACTACTAGTAGTAAGTAGTAATTTAAAGACTGCAGTAGAAG 715
 Db 1064 CACAGCTAAGTAGTTATAAAGCTAAATAATTAAGAGATTTAAAGATTTATGTAGATG 1123
 QY 716 AGTTACAAAATTAAGTCTAGTTATTTCTAATACTACAACTTTAGCTGGTGATGACAGAA 775
 Db 1124 ATTTAAACATATATATATCTTATTTCAANTGTTGTACAGTAGCAGGAGAGATAGAA 1183
 QY 776 TACAAACAGCTATAGAGATTAAGTAAGATAATTTACAAATATATGATGGCGAGAAATCAGATC 835
 Db 1184 TAGAACTGCTATAGAAATTAAGTAGTAATATTAATAATCTGAT-----GATA 1231
 QY 836 ATTCAGCTGATTTAAAGAGATTTAAAGATGTTTAAATGTTGTATTTAGTAGTGCAGATGCAGTAG 895
 Db 1232 AAAATGCAATAACTGATAAGCAGTTAATGATATAGTATTTAGTTGGATCTACATCTATAG 1291
 QY 896 TAGATGATTTAGTTGGGCTCTTTAGCAGCAGAAAGATGCTCCACTATTTAACTT 955
 Db 1292 TTGATGCTCTGTTGTCATCCATAGCTTCAGAAACAGAGCTCCATTTATTTAACTT 1351
 QY 956 CAAAAGATAAATTAGATTCGTCAGTAAATCTGAAATTAAGAGAGTTTGTAGACTTAAAA 1015
 Db 1352 CAAAAGATAAATTAGATTCATCAGTAAATCTGAAATTAAGAGAGTTTGAATCTTAAAGA 1411
 QY 1016 CTTCAACAGAGTA---ACAGGAAAAACAGTTTATATAGTCTGGTGGAGTTAATAGTGAT 1072
 Db 1412 GTGACACTGGTATAAATACTTCTAAAAAGTTTATTTAGTGGTGGAGTTAATTTCTATAT 1471
 QY 1073 CTAAGAGTTGTAACAGATTAAGATCAATGGGATTAAGATTTAAAGATTTCTCAGGTG 1132
 Db 1472 CTAAGATGTAGAAATGAATTTGAAACATGGGCTTTAAAGTTACTAGATTTACAGGAG 1531
 QY 1133 ATGATAGATATGAACCTTTTAAAAATAGCAGGTGAATPAGGCTTAGATAATGATAAGG 1192
 Db 1532 AAGACAGATCAGAACTCTTTAGCAATAGCTGATGAATAGGTCTTTGATAATGATAAAG 1591
 QY 1193 CTTATGATGTTGGTGGAAACAGGATAGCAGATGCCATGAGTATAGCTTCAGTTGCTTCTA 1252
 Db 1592 CATTTGATGTTGGTGGTACTGGATTTAGCAGATGCTATGATATAGCTCCAGTTGCTTCT- 1650
 QY 1253 CTAATTTAGATGGTAAATGGTGTGTTAGATAGAACAAATGGACATGCTACTCCAATAGTTG 1312
 Db 1651 -----CAACTTAAAGATGGAGATGCTCTCCAATAGTAG 1684
 QY 1313 TTGTAGATGGAAGAGCTGATAAATACTGATGACTTAGATAGTTTCTTAGGAAGCGCTG 1372
 Db 1685 TTGTAGATGGAAGAGCAAGAAATAGTGTGATGCTAGAGTTTCTTAGGAACCTTCTG 1744
 QY 1373 ATGTAGATATATAGTGGATTTGCCAAGTGATCTGAAAGAGATGGAAAGAGCTATATCAG 1432
 Db 1745 ATGTTGATATAATAGTGGAAAAAATAGCGTATCTAAAGAGATTTGAAGAGTCAATAGATA 1804
 QY 1433 ATGCTACTGGTAAAGGCTTACAGAGTTAAAGCGGACGATAGACAGACACTAACTCTG 1492
 Db 1805 GTGCAACTGGAAAACTCCAGATAGAAATAGTGGAGATGATAGACAGCACTAATGCTG 1864
 QY 1493 AAGTTATAAAAAATATTTATGCTAATGATCTGAAATAGCTAAAGCTGCAGTTTATAGATA 1552
 Db 1865 AAGTTTAAAGAGATGATTTTC----- 1890
 QY 1553 AAGATTCAGGTGCTTCAAGTAGTAGCAGAGATTTTAAATTTCTATGTAGCTTAAAGATG 1612
 Db 1891 -----ACAGATGGTGAAGTTGTGAATTTACTTTGTTGCAAAAGATG 1930
 QY 1613 GATCTACAAAAGAGATCAATTAGTTGATGCAATTAGCAGTAGGAGCTGTTGCTGG----- 1667

Db 1931 GTTCTACTAAAGAGATCAATTTAGTAGATGCTTTAGCAGCAGCAATAGCAGGTAGAT 1990
 QY 1668 -----ATATAAACTTCTCAGTTGTATTAGCTACTGATTTCTTTATCTTCTGATCAATCGG 1723
 Db 1991 TTAAGGAGTCTCCAGCTCAATCATCTAGCTACTGATCTTTATCTTCTGACCAAAATG 2050
 QY 1724 TTGCTATTAAGCAAAAGTTGTAGGAGAAAAATATTTCTAAAGATTTAAACAAGTTGGTCAAG 1783
 Db 2051 TAGCTGTAAGTAAAGCAGTTCTTAAAGATGGTGGAACTAACTTAGTTCAAGTAGGTAAAG 2110
 QY 1784 GAATAGCTTAATTCAGTTATTAACAACAAATGAAAGATTTATTAGATATG 1830
 Db 2111 GTATAGCTTCTTTCAGTTATTAACAACAAATGAAAGATTTATTAGATATG 2157

Search completed: October 6, 2005, 06:23:53
 Job time : 1048 secs

This Page Blank (uspto)